THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

CALIBRATION OF MULTIVARIATE GENERALIZED HYPERBOLIC DISTRIBUTIONS USING THE EM ALGORITHM, WITH APPLICATIONS IN RISK MANAGEMENT, PORTFOLIO OPTIMIZATION AND PORTFOLIO CREDIT RISK

By

WENBO HU

A Dissertation submitted to the
Department of Mathematics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2005

The members of the Committee approve the Dissertation of Wenbo Hu defended on Oct 28, 2005.

Alec Kercheval
Professor Directing Dissertation

Fred Huffer
Outside Committee Member

Bettye Case
Committee Member

Warren Nichols
Committee Member

Craig Nolder
Committee Member

The Office of Graduate Studies has verified and approved the above named committee members.

To my father Hu, Shaoren and mother He, Xihua

ACKNOWLEDGEMENTS

I would first like to thank my father Hu, Shaoren, mother He, Xihua and all my family members. Without their 18 years support, I can not go this far.

Dr. Kercheval guides me into the field of credit risk, risk management and portfolio optimization, provides tons of ideas, and offers infinite help and patience during my study in FSU. I own him a lot. I would like to acknowledge my gratitude to my committee members, Dr. Case, Dr. Huffer, Dr. Nichols, and Dr. Nolder who offer numerous help.

Many thanks for M. Galloway, B. Montin, J. Moreno, W. Zhu, A. Culham, J. Zhang, E. Yoo, and all other graduate fellows who give me some inspirations.

Special thanks go to K. Bell and R. Farr at Bell Trading. Without their help, this paper can not be finished. A lot of ideas in this paper are triggered by them. I also need to thank all my colleagues who provide comments and helps.

— Wenbo

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The distributions of many financial quantities are well-known to have heavy tails, exhibit skewness, and have other non-Gaussian characteristics. In this dissertation we study an especially promising family: the multivariate generalized hyperbolic distributions (GH). This family includes and generalizes the familiar Gaussian and Student t distributions, and the so-called skewed t distributions, among many others.

The primary obstacle to the applications of such distributions is the numerical difficulty of calibrating the distributional parameters to the data. In this dissertation we describe a way to stably calibrate GH distributions for a wider range of parameters than has previously been reported. In particular, we develop a version of the EM algorithm for calibrating GH distributions. This is a modification of methods proposed in McNeil, Frey, and Embrechts (2005), and generalizes the algorithm of Protassov (2004). Our algorithm extends the stability of the calibration procedure to a wide range of parameters, now including parameter values that maximize log-likelihood for our real market data sets. This allows for the first time certain GH distributions to be used in modeling contexts when previously they have been numerically intractable.

Our algorithm enables us to make new uses of GH distributions in three financial applications. First, we forecast univariate Value-at-Risk (VaR) for stock index returns, and we show in out-of-sample backtesting that the GH distributions outperform the Gaussian distribution. Second, we calculate an efficient frontier for equity portfolio optimization under the skewed-t distribution and using Expected Shortfall as the risk measure. Here, we show that the Gaussian efficient frontier is actually unreachable if returns are skewed t distributed. Third, we build an intensity-based model to price Basket Credit Default Swaps by calibrating the skewed t distribution directly, without the need to separately calibrate

the skewed t copula. To our knowledge this is the first use of the skewed t distribution in portfolio optimization and in portfolio credit risk.

CHAPTER 1

Introduction

1.1 Calibration of Generalized Hyperbolic Distributions

It has long been known that financial returns distributions are often not normally distributed, but the technical difficulties of dealing with non-normal distributions have often stood in the way of using them in financial modeling. In this dissertation, we address some of these problems in the context of the generalized hyperbolic (GH) distributions, a parametric family that has some useful properties and includes many familiar distributions such as the Student t, Gaussian, Cauchy, variance gamma, and others.

Barndorff-Nielsen (1977) introduced generalized hyperbolic distributions in the study of grains of sand. Later, Blæsild and Sørensen (1992) developed a computer program called HYP to fit hyperbolic distributions by using a maximum log-likelihood method. They reported it is impossible to calibrate the hyperbolic distribution if the dimension is greater or equal to four. Prause (1999) provided detailed derivations of the derivatives of the log-likelihood function for generalized hyperbolic distributions in his dissertation, and he applied HYP to calibrate the three dimensional generalized hyperbolic distributions. Eberlein and Keller(1995) fit the univariate hyperbolic distribution to return series of German equities and get a high accuracy fit. It is very time consuming to calibrate multivariate generalized hyperbolic distributions if we maximize the log-likelihood directly so that there have been very few applications of multivariate generalized hyperbolic distributions. Prause (1999) simplified the calculation by considering the much simpler symmetric generalized hyperbolic distributions. It is clear from this work that the calibration of generalized hyperbolic distributions has been difficult.

The EM algorithm is a powerful method for calibration that has been used for some time in various contexts. Dempster, Laird and Rubin(1977) showed that the EM(expecation maximization) algorithm can be used to find maximum likelihood estimates. Liu and Rubin(1995) used the EM algorithm and its extensions $ECME$(expectation conditional maximization either) and $MCECM$(multi cycle expectation conditional maximization) to calibrate the Student t distribution. The EM algorithm was used by Protassov(2004) to calibrate the λ fixed multivariate generalized hyperbolic distributions by maximizing the augmented log-likelihood. He also reported he is the first one who can calibrate multivariate generalized hyperbolic distributions when the dimension is greater than three. He fit the five dimensional normal inverse Gaussian(NIG) distribution to returns series on foreign exchange rates. Aas and Hobæk Haff(2005a and 2005b)and Aas, Dimakos and Hobæk Haff(2005c) used NIG and skewed t to model financial returns using EM algorithm.

All this work uses a particular parametrization of the GH family introduced by Barndorff-Nielsen (1977), in part because that is the parametrization implemented in the HYP program. It is consistent with the description of GH by Barndorff-Nielsen and Blæsid (1981) as normal mean variance mixtures (see chapter 2), where the mixing distribution is a generalized inverse Gaussian distribution(GIG). However, under this standard parametrization, the parameters of the underlying GIG distribution are not preserved by linear transformations.

An important innovation was made by McNeil, Frey and Embrechts(2005), where they gave a new parametrization of GH and the EM algorithm under this parametrization. Under this new parametrization, the linear transformation of generalized hyperbolic distributions remains in the same sub-family characterized by the the parameters of GIG.

We follow the framework of Liu and Rubin(1995), Protassov(2004) and McNeil, Frey and Embrechts(2005)'s EM algorithm and provide our own algorithm. Protassov's algorithm can be regarded as a special case of our algorithm. McNeil et al.(2005) mentioned that his algorithm might be unstable. Protassov's might be unstable too although he didn't mention this. Our algorithm solves the stability problem and we can calibrate the whole family of generalized hyperbolic distributions. Importantly, we provide a fast algorithm for some sub-families of GH such as normal inverse gaussian (NIG), variance gamma (VG), skewed t, and Student t etc.

1.2 Applications

We apply generalized hyperbolic distributions in three financial applications: risk management, portfolio optimization and pricing of portfolio credit risk, as we now describe.

Risk management. Value at Risk (VaR) based on the normal distribution has been considered as the standard risk measure since J.P. Morgan launched RiskMetrics in 1994. In 1995, the Basel Committee on Banking Supervision suggested using the 10 day VaR at the 99% level. Yet, the financial return series is heavy tailed and leptokurtic; large losses occur far away from the VaR based on a normal distribution. For financial return series, at the 99% level, VaR based on a normal distribution usually underestimates the true risk.

Risk managers can not neglect this problem. VaR_α depends only on the choice of underlying distribution. Semi heavy tailed GH has been considered as an alternative. Barbachan, Farias and Ornelas(2003) and Fajardo and Farias (2003) applied univariate generalized hyperbolic distributions (with the usual parametrization) to model the Brazilian data to get more accurate VaR measurements. We use the new parametrization of generalized hyperbolic distributions to fit the univariate return series and calculate the corresponding VaR_α.

Typical the financial return series usually is not independently and identically distributed (*i.i.d.*), so before modeling returns with a particular choice of distribution, we need to filter the data. We use a $GARCH(1,1)$ filter with Student t or Gaussian innovations to filter our data.

After we get the *i.i.d.* return series, we calibrate NIG, VG, skewed t and hyperbolic distributions chosen from the general GH family and then implement a VaR backtesting procedure suggested by McNeil(1999). We show that use of GH distributions leads to better VaR forecasts than Gaussian.

Portfolio optimization. Portfolio optimization is based on trading off risk and return. The construction of an efficient frontier – portfolios with minimum risk for a given return – depends on two inputs: the choice of risk measure, and the probability distribution used to model returns.

For this purpose one needs to employ some precise concept of “risk” and a accurate description of the distribution of returns.

Markowitz (1952) suggested using the standard deviation of portfolio return as a risk measure, and, thinking of returns as normally distributed, described the efficient frontier, which is composed of fully invested portfolios having minimum risk for a given specified return. However, using standard deviation as the risk measure has the drawback that it is generally insensitive to extreme events. VaR can describe more about extreme events, but it can not aggregate risk in the sense of being subadditive on portfolios. This is a well-known difficulty that is addressed by the concept of a "coherent risk measure" in the sense of Artzner, Delbaen, Eber, and Heath(1999). A popular example of a coherent risk measure is the expected shortfall (ES).

It turns out, by a result of Embrechts, McNeil, and Straumann (2001), that when the underlying distribution is Gaussian – or more generally any "elliptical" distribution – no matter what positive homogeneous and translation invariant risk measure, no matter what confidence level, the optimized portfolio composition given a certain return will be the same as the traditional Markowitz style portfolio composition. Only the choice of distribution will affect the optimized portfolio. As mentioned before, portfolio managers can not neglect the deviation of financial returns series from a multivariate normal distribution. Other heavy tailed elliptical distributions, such as Student t and symmetric generalized hyperbolic distributions, and non-elliptical distributions, such as the skewed t distribution, can be used to model financial returns series.

Rockafellar and Uryasev (1999) showed that the minimization of ES does not require knowing VaR first, and construct a new objective function. By minimizing this new objective function, we can get VaR and ES. We carry this out and use Monte Carlo simulation to approximate that new objective function by sampling the multivariate distributions. This allows us to construct efficient frontiers for a variety of distributions. One consequence we describe is that the usual Gaussian efficient frontier is actually unreachable if we believe returns are Student t or skewed t distributed.

Correlations, copulas, and Credit Risk. Copulas have become a popular way to describe and construct the dependence structure of multivariate distributions. Credit events tend to happen together because of connections in business. Corresponding to the heavy tail property in univariate distributions, tail dependence is used to model the co-occurrence of extreme events. The Student t copula is tail dependent while the Gaussian copula is tail independent.

The calibration of a Student t copula is separate from the calibration of marginal distributions. It is generally suggested to use the empirical distributions to fit the margins.

In the pricing of multiname credit derivatives such as basket credit default swaps (CDS), and collateralized debt obligations (CDO), the most important issue is the correlations between those default obligors. Copulas can be introduced to model these correlations by using the correlations of corresponding equity prices. We show that Kendall's τ correlation remains invariant under monotone transformations. This is the foundation of modeling the correlation of credit events by using the correlation of underlying equities via copulas, though nobody mentions this correlation invariance property.

In fact, we can also use a multivariate distribution of underlying equity prices instead of using copulas to model the correlation of credit events with the correlation of underlying equities by using this correlation invariance property. The only difference between a copula approach and a distribution approach lies in the calibration procedure. For a copula approach, the calibrations of marginal distributions and copula are separate, while for a distribution approach, the two are jointly calibrated.

For the Student t copula, there is still no good method to calibrate the degree of freedom ν except to find it by direct search. The calibration of Student t copula takes days while the calibration of skewed t or Student t distribution takes minutes.

We use the Student t and skewed t distributions to model the correlations of default obligors in the final chapter.

CHAPTER 2

Calibration of Multivariate Generalized Hyperbolic Distributions

Generalized hyperbolic distributions can be parameterized in several ways. Most of the literature defines the density of generalized hyperbolic distributions directly, such as Prause(1999) and Protassov(2004). However, such definition makes the application of multivariate generalized hyperbolic distributions inconvenient since some important characterizing parameters are not invariant under linear transformations. We follow McNeil , Frey, and Embrechts(2005) to introduce the generalized hyperbolic distributions from the definition of generalized inverse Gaussian distributions by mean-variance mixing method. Under this parametrization, the linear transformations of generalized hyperbolic random vectors remain in the same sub family of generalized hyperbolic distributions which is characterized by some type parameters.

We introduce the building block, generalized inverse Gaussian distributions, in Section 2.1 and generalized hyperbolic distributions in Section 2.2. Some univariate sub families of generalized hyperbolic distributions have been applied successfully in the modeling of financial returns, which can be found in Eberlein and Keller(1995), and Barndorff-Nielsen(1997) etc. We discuss some sub families in Section 2.2 after we introduce the generalized hyperbolic distributions. In Section 2.3, we discuss the one dimensional generalized hyperbolic distributions and their tail behavior. We try to explain the role that the parameters of generalized hyperbolic distributions play in the modeling of the distributions. In Section 2.4, we give detailed EM algorithms for the multivariate generalized hyperbolic distributions and all their well known limiting distributions by following the EM algorithm in McNeil , Frey, and Embrechts(2005). In Section 2.5, we calibrate real financial returns series for generalized hyperbolic distributions and the limiting or special distributions using McNeil et al(2005), Protassov(2004) and our EM algorithm and compare the performance.

2.1 The Generalized Inverse Gaussian Distributions

Definition 2.1.1 Gamma Distribution. *The random variable X is said to have a* gamma distribution*, written as $X \sim Gamma(\alpha, \beta)$, if its probability density function is*

$$f(x) = \beta^{\alpha} x^{\alpha-1} e^{-\beta x}/\Gamma(\alpha), \quad x > 0, \quad \alpha > 0, \quad \beta > 0. \tag{2.1.1}$$

We have the following formulas:

$$E(X) = \frac{\alpha}{\beta}, \tag{2.1.2}$$

$$Var(X) = \frac{\alpha}{\beta^2}, \tag{2.1.3}$$

$$E(\log(x)) = \psi(\alpha) - \log(\beta), \tag{2.1.4}$$

where $\psi(x) = d\log(\Gamma(x))/dx$ is digamma function. When $X \sim Gamma(\nu/2, 1/2)$, we say X is χ^2_v distributed *.

Definition 2.1.2 Inverse Gamma Distribution. *The random variable X is said to have an* inverse gamma distribution*, written as $X \sim InverseGamma(\alpha, \beta)$, if its probability density function is*

$$f(x) = \beta^{\alpha} x^{-\alpha-1} e^{-\beta/x}/\Gamma(\alpha), \quad x > 0, \quad \alpha > 0, \quad \beta > 0. \tag{2.1.5}$$

If $X \sim Gamma(\alpha, \beta)$, then $1/X \sim InverseGamma(\alpha, \beta)$.

We have following formulas which are needed later,

$$E(X) = \frac{\beta}{\alpha - 1}, \quad \text{if } \ \alpha > 1 \tag{2.1.6}$$

$$Var(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad \text{if } \ \alpha > 2 \tag{2.1.7}$$

$$E(\log(x)) = \log(\beta) - \psi(\alpha). \tag{2.1.8}$$

*In some papers and Matlab, a gamma distributed random variable with parameters (a, b), is defined to be $X \sim Gamma(a, 1/b)$, where $a = \alpha, b = 1/\beta$.

Definition 2.1.3 Modified Bessel Function of the Third Kind with Index λ. *The* integral presentation of the modified Bessel function of the third kind with index λ *can be found in Barndorff-Nielsen et al. (1981),*

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1} e^{-\frac{x}{2}(y+y^{-1})} dy, \quad x > 0. \tag{2.1.9}$$

Definition 2.1.4 Generalized Inverse Gaussian distribution(GIG**)**. *The random variable* X *is said to have a* generalized inverse gaussian(GIG) *distribution if its probability density function is*

$$h(x;\lambda,\chi,\psi) = \frac{\chi^{-\lambda}(\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right), x > 0, \tag{2.1.10}$$

where K_λ *is a modified Bessel function of the third kind with index* λ *and the parameters satisfy*

$$\begin{cases} \chi > 0, \psi \geq 0 & if \lambda < 0 \\ \chi > 0, \psi > 0 & if \lambda = 0 \\ \chi \geq 0, \psi > 0 & if \lambda > 0 \end{cases}$$

In short, we write $X \sim N^-(\lambda,\chi,\psi)$ *if* X *is* GIG *distributed.*

The following formulas for GIG distributed variable X when $\chi > 0$ and $\psi > 0$ may be needed later,

$$E(X^\alpha) = \left(\frac{\chi}{\psi}\right)^{\alpha/2} \frac{K_{\lambda+\alpha}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}, \tag{2.1.11}$$

especially when $\alpha = \pm\, 1$ and 2, and

$$E(\log(X)) = \frac{dE(X^\alpha)}{d\alpha}|_{\alpha=0}, \tag{2.1.12}$$

where equation 2.1.12 needs to be evaluated numerically. More details of GIG can be found in Jørgensen(1982) and Barndorff-Nielsen and Stelzer(2004).

The generalized hyperbolic distributions can be introduced from GIG distributions and their properties are greatly characterized by the corresponding GIG distributions. Therefore, it is meaningful to discuss some special cases of GIG distributions.

When $\lambda = -0.5$, GIG becomes the inverse Gaussian distribution and its corresponding generalized hyperbolic distribution is the normal inverse Gaussian distribution. In addition, we have an explicit form of the Bessel function,

$$K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2}} x^{-1/2} e^{-x}. \tag{2.1.13}$$

When $\chi = 0$ and $\lambda > 0$, GIG becomes the so-called gamma distribution and it is denoted by $X \sim Gamma(\lambda, \psi/2)$. This limiting case of GIG will lead to the limiting case of generalized hyperbolic distributions: the variance gamma distribution. The following asymptotic formula is useful when calculating the limiting density of GIG and GH,

$$K_\lambda(x) \sim \Gamma(\lambda) 2^{\lambda-1} x^{-\lambda} \quad as \quad x \downarrow 0. \tag{2.1.14}$$

When $\psi = 0$ and $\lambda < 0$, GIG becomes the so-called inverse gamma distribution and it is denoted by $X \sim InverseGamma(-\lambda, \chi/2)$. This limiting case of GIG will lead to a limiting case of generalized hyperbolic distributions called the skewed t distribution, under some conditions mentioned in Section 2.2.2. The following asymptotic formula is useful when calculating the limiting density of GIG and GH,

$$K_\lambda(x) \sim \Gamma(-\lambda) 2^{-\lambda-1} x^{\lambda} \quad as \quad x \downarrow 0. \tag{2.1.15}$$

The following fact of Bessel function is also very useful,

$$K_\lambda(x) = K_{-\lambda}(x). \tag{2.1.16}$$

More details about the limiting case of GIG can be found in Eberlein and Hammerstein(2003).

2.2 The Multivariate Generalized Hyperbolic Distributions

Generalized hyperbolic distributions can be represented as a normal mean-variance mixture where the mixture variable is GIG distributed.

Definition 2.2.1 Normal Mean-Variance Mixture. *The random variable* $\mathbf{X}$ *is said to have a* multivariate normal mean-variance mixture distribution *if*

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W} A \mathbf{Z}, \quad where \tag{2.2.1}$$

1. $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I_k})$,

2. $W \geq 0$ *is a positive, scalar-valued r.v. which is independent of* $\mathbf{Z}$,

3. $A \in \mathbb{R}^{d \times k}$ *is a matrix,*

4. $\boldsymbol{\mu}$ *and* $\boldsymbol{\gamma}$ *are parameter vectors in* $\mathbb{R}^d$*.*

From the definition, we can see that

$$\mathbf{X} \mid W \sim N_d(\boldsymbol{\mu} + W\boldsymbol{\gamma}, W\Sigma), \tag{2.2.2}$$

where $\Sigma = AA'$. This is also why it is called normal mean-variance mixture distribution. We can get the following moment formulas easily from the mixture definition

$$E(\mathbf{X}) = \boldsymbol{\mu} + E(W)\boldsymbol{\gamma}, \tag{2.2.3}$$

$$COV(\mathbf{X}) = E(W)\Sigma + var(W)\boldsymbol{\gamma}\boldsymbol{\gamma}', \tag{2.2.4}$$

when the mixture variable W has finite variance $var(W)$.

The mixture variable W can be explained as a shock which changes the volatility and mean of the normal distribution in economics.

If the mixture variable W is GIG distributed, then $\mathbf{X}$ is said to have a generalized hyperbolic distribution.

If $\boldsymbol{\gamma} = \mathbf{0}$, then $\mathbf{X}$ is said to have a symmetric generalized hyperbolic distribution and Schmidt(2003a) showed that it is also an elliptical distribution; otherwise, it is not elliptical. More details of elliptical distributions can be found in Fang, Kotz and Ng(1990).

Theorem 2.2.2 Generalized Hyperbolic Distributions(GH)(McNeil , Frey, and Embrechts(2005)). *If the mixing variable* $W \sim N^-(\lambda, \chi, \psi)$*, then the joint density of* d-dimensional generalized hyperbolic distributions *in the non-singular case (*Σ *has rank* d*) is given by,*

$$f(\boldsymbol{x}) = c\frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\chi + (\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\,(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right)e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{(\chi + (\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\,(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right)^{\frac{d}{2}-\lambda}}, \tag{2.2.5}$$

where the normalizing constant is

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{\lambda}(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}K_{\lambda}(\sqrt{\chi\psi})},$$

and $|\cdot|$ *denotes the determinant.*

Proof: From the definition of normal mean-variance mixture distribution, the joint density of $\mathbf{X}$ is given by

$$f(\boldsymbol{x}) = \int_0^\infty \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}w^{\frac{d}{2}}} exp\left\{-\frac{(\boldsymbol{x}-\boldsymbol{\mu}-\boldsymbol{w\gamma})'(w\Sigma)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}-\boldsymbol{w\gamma})}{2}\right\} h(w)dw,$$

where $h(w)$ is the density of W. $f(\boldsymbol{x})$ can be rewritten as,

$$f(\boldsymbol{x}) = \int_0^\infty \frac{e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}w^{\frac{d}{2}}} exp\left\{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2w} - \frac{\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}}{2/w}\right\} h(w)dw.$$

From equation 2.1.10, after some rearrangements, we can get the following,

$$f(\boldsymbol{x}) = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{\lambda}e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}K_\lambda(\sqrt{\chi\psi})}\times$$

$$\frac{1}{2}\int_0^\infty w^{\lambda-\frac{d}{2}-1} exp\left\{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})+\chi}{2w} - \frac{\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}+\psi}{2/w}\right\} dw.$$

By setting,

$$y = w\frac{\sqrt{(\psi+\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}}{\sqrt{(\chi+(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))}},$$

after some rearrangements, we can get,

$$f(\boldsymbol{x}) = c\frac{e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{(\chi+(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))(\psi+\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right)^{\frac{d}{2}-\lambda}}\times$$

$$\frac{1}{2}\int_0^\infty y^{\lambda-\frac{d}{2}-1} exp\left\{-\frac{1}{2}\sqrt{(\chi+(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))(\psi+\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\left[\frac{1}{y}+y\right]\right\} dy$$

By definition 2.1.3, we can get the density of generalized hyperbolic distributions. ∎

In the sequel, we denote the quadratic form by

$$\rho_x = (\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \tag{2.2.6}$$

By equation 2.1.11, equation 2.2.3 and equation 2.2.4, the mean and covariance of GH distributed random vector $\mathbf{X}$ are given by

$$E(\mathbf{X}) = \boldsymbol{\mu} + \boldsymbol{\gamma}\left(\frac{\chi}{\psi}\right)^{1/2}\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}, \tag{2.2.7}$$

$$COV(\mathbf{X}) = \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}\Sigma + \boldsymbol{\gamma}\boldsymbol{\gamma}' \left(\left(\frac{\chi}{\psi}\right) \frac{K_{\lambda+2}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} - \left(\left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} \right)^2 \right), \tag{2.2.8}$$

when the mixing variable W has finite variance.

For a d-dimensional normal distributed random vector, $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, it is well known that the characteristic function is

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(e^{i\boldsymbol{t}'\boldsymbol{X}}) = e^{i\boldsymbol{t}'\boldsymbol{\mu} - \boldsymbol{t}'\Sigma\boldsymbol{t}/2}. \tag{2.2.9}$$

From the mean-variance mixture definition, we can get the characteristic function of the generalized hyperbolic distributed random vector $\mathbf{X}$,

$$\begin{aligned} \phi_{\boldsymbol{X}}(\boldsymbol{t}) &= E\left(E(e^{i\boldsymbol{t}'\boldsymbol{X}}|W)\right) = E(e^{i\boldsymbol{t}'(\boldsymbol{\mu}+W\boldsymbol{\gamma}) - W\boldsymbol{t}'\Sigma\boldsymbol{t}/2}) \\ &= e^{i\boldsymbol{t}'\boldsymbol{\mu}}\hat{H}(\boldsymbol{t}'\Sigma\boldsymbol{t}/2 - i\boldsymbol{t}'\boldsymbol{\gamma}), \end{aligned} \tag{2.2.10}$$

where $\hat{H}(\theta) = E(e^{-\theta W})$ is the Laplace transform of the GIG distribution. The form of the characteristic function of the generalized hyperbolic distributions shows that the parameters coming from GIG play important roles in modeling GH. We use the notation $X \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$. The first three parameters are from the mixing distribution, while the last three are location, scale, and skewness parameters.

The characteristic function can be used to show that generalized hyperbolic distributions are closed under linear transformations.

Proposition 2.2.3 Linear Transformations of Generalized Hyperbolic Distributions(McNeil , Frey, and Embrechts(2005)). *If* $\boldsymbol{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ *and* $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$ *where* $B \in \mathbb{R}^{k\times d}$ *and* $\boldsymbol{b} \in \mathbb{R}^k$*, then* $\boldsymbol{Y} \sim GH_k(\lambda, \chi, \psi, B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B', B\boldsymbol{\gamma})$.

Proof:

$$\begin{aligned} \phi_{\boldsymbol{Y}}(\boldsymbol{t}) &= E(e^{i\boldsymbol{t}'(B\boldsymbol{X}+\boldsymbol{b})}) = e^{i\boldsymbol{t}'\boldsymbol{b}}\phi_{\boldsymbol{X}}(B'\boldsymbol{t}) \\ &= e^{i\boldsymbol{t}'(B\boldsymbol{\mu}+\boldsymbol{b})}\hat{H}(\boldsymbol{t}'B\Sigma B'\boldsymbol{t}/2 - i\boldsymbol{t}'B\boldsymbol{\gamma}). \end{aligned}$$

∎

This proposition shows that the linear transformations of generalized hyperbolic distribution still remain in the generalized hyperbolic distribution class generated by the same generalized inverse Gaussian distribution $N^-(\lambda, \chi, \psi)$ and this property is very useful in the portfolio management.

Corollary 2.2.4 Weighted Sum of Generalized Hyperbolic Random Variables. *If $B = \boldsymbol{\omega}^T = (\omega_1, \cdots, \omega_d)$, and $\mathbf{b} = \mathbf{0}$, then the portfolio $y = \boldsymbol{\omega}^T\mathbf{X}$ is a one dimensional generalized hyperbolic distribution, and*

$$y \sim GH_1(\lambda, \chi, \psi, \boldsymbol{\omega}^T\boldsymbol{\mu}, \boldsymbol{\omega}^T\Sigma\boldsymbol{\omega}, \boldsymbol{\omega}^T\boldsymbol{\gamma}). \tag{2.2.11}$$

Gaussian distributions and Student t distributions are closed under linear transformations. Generalized hyperbolic distributions are also closed under linear transformations. This corollary shows that the method used in portfolio risk management based on Gaussian or Student t distribution can also be used in generalized hyperbolic distributions. Gaussian distributions and Student t distributions are elliptical, while Generalized hyperbolic distributions are not, so that they can capture more characteristics about financial data series such as asymmetry but do not incur more difficulty in application. We can calculate the portfolio risk such as standard deviation, VaR, and ES etc. without Monte Carlo simulation. This corollary also shows that the marginal distributions are automatically obtained once we have calibrated the multivariate generalized hyperbolic distributions, i.e., $X_i \sim GH_1(\lambda, \chi, \psi, \mu_i, \Sigma_{ii}, \gamma_i)$.

2.2.1 Parameterizations

We can see from equation 2.2.5 that for any $c > 0$,

$$GH_d(\lambda, \chi/c, c\psi, \boldsymbol{\mu}, c\Sigma, c\boldsymbol{\gamma}) = GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}). \tag{2.2.12}$$

This identification problem does cause a lot of trouble in the calibration, where the redundant information causes the algorithm to be unstable. In practice, many people set the determinant of Σ to be 1 when fitting the data and call this new dispersion matrix Δ.

McNeil , Frey, and Embrechts(2005) mentioned that after rescaling, we can obtain the usual definition of generalized hyperbolic distributions by following the transformation

$$\boldsymbol{\beta} = \Delta^{-1}\boldsymbol{\gamma},\ \delta = \sqrt{\chi},\ \alpha = \sqrt{\psi + \boldsymbol{\beta}'\Delta\boldsymbol{\beta}}. \tag{2.2.13}$$

These parameters must satisfy the constraints:

$$\begin{cases} \delta > 0, \alpha^2 \geq \boldsymbol{\beta}'\Delta\boldsymbol{\beta} & if \lambda < 0 \\ \delta > 0, \alpha^2 > \boldsymbol{\beta}'\Delta\boldsymbol{\beta} & if \lambda = 0 \\ \delta \geq 0, \alpha^2 > \boldsymbol{\beta}'\Delta\boldsymbol{\beta} & if \lambda > 0. \end{cases}$$

We will discuss more about the equivalence in one dimension in Section 2.3. Blæsild(1981) used this parametrization to show that generalized hyperbolic distributions are closed under linear operations and conditioning; however, unfortunately, α, and δ are not generally invariant. More details can be found in Prause(1999).

McNeil , Frey, and Embrechts (2005) set the determinant of Σ to be the determinant of the sample covariance matrix. From our calibration experience, the most stable way is to set χ to be a constant if $\lambda > -1$ and to set ψ to be a constant if $\lambda < 1$. See Section 2.4 for the difference of the algorithms.

2.2.2 Some Special Cases

Hyperbolic Distributions:

If $\lambda = 1$, we get the multivariate generalized hyperbolic distribution whose univariate margins are one-dimensional hyperbolic distributions.

If $\lambda = (d+1)/2$, we get the d-dimensional hyperbolic distribution, however, its marginal distributions are not hyperbolic distributions any more.

The one dimensional hyperbolic distribution is widely used in the modeling of univariate financial data, such as Eberlein and Keller(1995), Fajardo and Farias(2003).

Normal Inverse Gaussian Distributions(NIG):

If $\lambda = -1/2$, then the distribution is known as normal inverse Gaussian (NIG). NIG is largely used in the modeling of univariate financial return too. For NIG distribution, we have a fast algorithm to calibrate it. See Section 2.4 for the details of the algorithm.

Variance Gamma Distribution(VG):

If $\lambda > 0$ and $\chi = 0$, then we get a limiting case known as the variance gamma distribution. For the variance gamma distribution, we can calibrate all the parameters including λ. See Section 2.4 for the details of the algorithm.

By replacing χ by 0 and $\frac{\chi^{-\lambda/2}}{K_\lambda(\sqrt{\chi\psi})}$ by $\frac{\psi^{\lambda/2}}{\Gamma(\lambda)2^{\lambda-1}}$ in equation(2.2.5), we can get the density function of variance gamma.

The mean and covariance of a variance gamma distributed random vector $\mathbf{X}$ are

$$E(\mathbf{X}) = \boldsymbol{\mu} + \boldsymbol{\gamma}\frac{2\lambda}{\psi}, \tag{2.2.14}$$

$$COV(\mathbf{X}) = \frac{2\lambda}{\psi}\Sigma + \boldsymbol{\gamma}\boldsymbol{\gamma}'\frac{4\lambda}{\psi^2}. \tag{2.2.15}$$

Skewed t Distribution:

If $\lambda = -\nu/2, \chi = \nu$ and $\psi = 0$, we get a limiting case which is called the skewed t distribution by Demarta and McNeil(2005). The Student t distribution is widely used in the modeling of univariate financial data since we can model the heaviness of the tail by controlling the degree of freedom, ν. It can be used in the modeling of multivariate financial data too since the EM algorithm can be used to calibrate it. It is also widely used to model dependence by creating a Student t copula from the Student t distribution. Student t copulas are widely used in the modeling of financial correlations since they are upper tail and lower tail dependent and very easy to calibrate. However, they are symmetric and bivariate exchangeable. It is arguable that one should not use symmetric and exchangeable copulas since financial events tend to crash together more than boom together. Schmidt(2003a) showed that all copulas created from symmetric generalized hyperbolic distributions are tail independent. A Student t distribution is a limiting case of symmetric generalized hyperbolic distributions. It may be interesting to create the skewed t copula from skewed t distributions. More discussions about symmetry and exchangeability of Student t copula and its extension, skewed t copula, can be found in Demarta and McNeil(2005) and McNeil, Frey and Embrechts(2005).

By replacing ψ by 0, χ by ν, λ by $-\nu/2$ and $\frac{\psi^{\lambda/2}}{K_\lambda(\sqrt{\chi\psi})}$ by $\frac{\nu^{\nu/4}}{\Gamma(\nu/2)2^{\nu/2-1}}$ in equation(2.2.5), we can get the density function of skewed t distribution.

Definition 2.2.5 Skewed t Distribution. *If $\mathbf{X}$ is skewed t distributed, then the joint density function is given by*

$$f(\boldsymbol{x}) = c\frac{K_{\frac{\nu+d}{2}}\left(\sqrt{(\nu+\rho_x)\left(\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}\right)}\right)e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{(\nu+\rho_x)\left(\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}\right)}\right)^{-\frac{\nu+d}{2}}(1+\frac{\rho_x}{\nu})^{\frac{\nu+d}{2}}}, \tag{2.2.16}$$

where the normalizing constant is

$$c = \frac{2^{1-\frac{\nu+d}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}.$$

The mean and covariance of skewed t distributed random vector $\mathbf{X}$ are

$$E(\mathbf{X}) = \boldsymbol{\mu} + \boldsymbol{\gamma}\frac{\nu}{\nu-2}, \tag{2.2.17}$$

$$COV(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma + \boldsymbol{\gamma}\boldsymbol{\gamma}'\frac{2\nu^2}{(\nu-2)^2(\nu-4)}, \tag{2.2.18}$$

where the covariance matrix is only defined when $\nu > 4$.

Furthermore, if $\boldsymbol{\gamma} = \mathbf{0}$, using equation 2.1.14, we get the Student t distribution.

Definition 2.2.6 Student t Distribution. *If $\mathbf{X}$ is t distributed, then the joint density function is given by*

$$f(\boldsymbol{x}) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}(1 + \frac{\boldsymbol{\rho_x}}{\nu})^{-\frac{\nu+d}{2}}. \tag{2.2.19}$$

The mean and covariance of Student t distributed random vector $\mathbf{X}$ are

$$E(\mathbf{X}) = \boldsymbol{\mu}, \tag{2.2.20}$$

$$COV(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma, \tag{2.2.21}$$

where the covariance matrix is only defined when $\nu > 2$.

2.3 One Dimensional Generalized Hyperbolic Distributions

One dimensional generalized hyperbolic distributions, especially, the hyperbolic distribution and normal inverse Gaussian distribution, are well known for their better fit to univariate financial data. They provide heavy, or at least semi-heavy tailed distribution. In equation 2.2.5, setting $d = 1$ and $\Sigma = \sigma^2$, we can get the one dimensional generalized hyperbolic distributions.

Definition 2.3.1 Density Function of One Dimensional Generalized Hyperbolic Distributions($1dGH$). *If X is* one-dimensional generalized hyperbolic *distributed, then the density function is given by*

$$f(x;\lambda,\chi,\psi,\mu,\sigma,\gamma) = \frac{(\sqrt{\psi\chi})^{-\lambda}\psi^{\lambda}(\psi+\frac{\gamma^2}{\sigma^2})^{0.5-\lambda}}{\sqrt{2\pi}\sigma K_\lambda(\sqrt{\psi\chi})} \times \frac{K_{\lambda-0.5}\left(\sqrt{\left(\chi+\frac{(x-\mu)^2}{\sigma^2}\right)\left(\psi+\frac{\gamma^2}{\sigma^2}\right)}\right)e^{\frac{\gamma(x-\mu)}{\sigma^2}}}{\left(\sqrt{\left(\chi+\frac{(x-\mu)^2}{\sigma^2}\right)\left(\psi+\frac{\gamma^2}{\sigma^2}\right)}\right)^{0.5-\lambda}} \tag{2.3.1}$$

where K_λ denotes a modified Bessel function of the third kind with index λ, and $x \in \mathbb{R}$. The domain of variations of the parameters are $\mu \in \mathbb{R}$ and

$$\begin{cases} \chi > 0, \psi \geq 0 & if \lambda < 0 \\ \chi > 0, \psi > 0 & if \lambda = 0 \\ \chi \geq 0, \psi > 0 & if \lambda > 0. \end{cases}$$

If we set $\tilde{\gamma} = \frac{\gamma}{\sigma}$, then we can see that σ is the scale parameter and μ is the location parameter since only the term$\frac{x-\mu}{\sigma}$ and not x alone appears in the density function.

By setting,

$$c = \frac{1}{\sigma^2},\ \beta = \frac{\gamma}{\sigma^2},\ \delta = \sqrt{\frac{\chi}{c}},\ \alpha = \sqrt{\frac{\psi}{\sigma^2}+\beta^2},$$

we can get the following definition used in most of the literature by some algebra.

Theorem 2.3.2 Density Function of 1-d GH Distributions Used in Most Literature. *If X is one-dimensional generalized hyperbolic distributed, then the density function is given by*

$$gh(x;\lambda,\alpha,\beta,\delta,\mu) = a(\lambda,\alpha,\beta,\delta)\left(\delta^2+(x-\mu)^2\right)^{(\lambda-0.5)/2} \times K_{\lambda-0.5}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) exp\left(\beta(x-\mu)\right), \tag{2.3.2}$$

and the normalizing constant is,

$$a(\lambda,\alpha,\beta,\delta) = \frac{(\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-0.5}\delta^{\lambda}K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}.$$

The domain of variations of the parameters are $\mu \in \mathbb{R}$ and

$$\begin{cases} \delta > 0, |\beta| \leq \alpha & if \lambda < 0 \\ \delta > 0, |\beta| < \alpha & if \lambda = 0 \\ \delta \geq 0, |\beta| < \alpha & if \lambda > 0. \end{cases}$$

By setting $\lambda = -1/2$ in equation 2.3.1, we get the one-dimensional normal inverse Gaussian distribution.

Definition 2.3.3 Density Function of NIG. *If X is* one-dimensional *NIG distributed, then the density function is given by*

$$f(x;\lambda,\chi,\psi,\mu,\sigma,\gamma) = \frac{\chi^{1/2}(\psi+\frac{\gamma^2}{\sigma^2})}{\pi\sigma e^{-\sqrt{\chi\psi}}} \times \frac{K_1\left(\sqrt{\left(\chi+\frac{(x-\mu)^2}{\sigma^2}\right)\left(\psi+\frac{\gamma^2}{\sigma^2}\right)}\right)e^{\frac{\gamma(x-\mu)}{\sigma^2}}}{\left(\sqrt{\left(\chi+\frac{(x-\mu)^2}{\sigma^2}\right)\left(\psi+\frac{\gamma^2}{\sigma^2}\right)}\right)} \tag{2.3.3}$$

By setting $\lambda = -1/2$, $\gamma = 0$, and $\psi \to 0$, we get the one-dimensional Cauchy distribution.

Definition 2.3.4 Density Function of Cauchy Distribution. *If X is* one-dimensional Cauchy(μ,δ) *distributed, then the density function is given by*

$$f(x) = \frac{\delta}{\pi(\delta^2+(x-\mu)^2)}, \tag{2.3.4}$$

where $\delta = \sigma\sqrt{\chi}$.

The stable distribution[†] is well known for its heavy tail and is widely used in the modeling of univariate financial data and portfolio management. Theoretical framework can be found in Nolan(2005) and financial applications can be found in Rachev and Mittnik(2000).

Definition 2.3.5 Stable Distribution. *A random variable X is said to have a* stable distribution *if there are parameters $\alpha \in (0,2], a \in [0,\infty), \beta \in [-1,1]$ and $b \in \mathbb{R}$ such that its characteristic function has the following form:*

$$\Psi_X(t) = \begin{cases} exp\left\{-a^\alpha|t|^\alpha\left(1-i\beta(sign\ t)\tan\frac{\pi\alpha}{2}+ibt\right)\right\} & if\ \alpha \neq 1 \\ exp\left\{-a|t|\left(1+i\beta\frac{2}{\pi}(sign\ t)\ln|t|\right)+ibt\right\} & if\ \alpha = 1 \end{cases}$$

where α is the index of stability, a is scale parameter, b is location parameter and β is skewness parameter. If β and μ are zero, X is said to be symmetric stable, which is denoted $X \sim S\alpha S$.

[†]The calibration of stable distribution can be found in http://pws.prserv.net/jpnolan/

A Gaussian distribution $N(\mu, \sigma^2)$ is stable with($\alpha = 2, \beta = 0, a = \sigma/\sqrt{2}$ and $b = \mu$). A Cauchy (μ, δ) distribution is stable with($\alpha = 1, \beta = 0, a = \delta$ and $b = \mu$).

However, a common criticism of the stable distribution, such by as Eberlein et al.(1995), is that their tails are too heavy.

Tails of Some Distributions:

1. **Tails of Cauchy Distribution**. If X is of Cauchy distribution, then

$$f(x) \sim \frac{c}{x^2} \quad as \quad x \to \pm\infty, \tag{2.3.5}$$

where c is a constant. Cauchy distribution has neither mean nor variance.

2. **Tail of Stable Distribution(Nolan(2005))**. If X is of stable distribution, then

$$f(x) \sim cx^{-\alpha-1} \; as \; x \to \infty, \tag{2.3.6}$$

which means that the tail is a power function. For $0 < \alpha < 2$, $E|X|^p < \infty$ if $0 < p < \alpha$. In particular $EX^2 = \infty$ for all non-Gaussian stable distributions, which means they have infinity second moment.

3. **Tails of Gaussian Distribution.** If X is of Gaussian distribution, then

$$f(x) \sim ce^{-x^2/2} \; as \; x \to \pm\infty, \tag{2.3.7}$$

which means that the tail is an exponential function of square function. Both tails are very thin.

4. **Tails of Student t Distribution.** If X is of Student t distribution, then

$$f(x) \sim c|x|^{-\nu-1} \; as \; x \to \pm\infty, \tag{2.3.8}$$

which means that the tail is a power function.

From equation 2.3.1, we can get the tails of generalized hyperbolic distributions.

Theorem 2.3.6 Tails of Generalized Hyperbolic Distributions. *If X is of generalized hyperbolic distributions, then*

$$f_{GH}(x; \lambda, \chi, \psi, \mu, \sigma, \gamma) \sim c|x|^{\lambda-1}e^{-\alpha|x|+\beta x} \; as \; x \to \pm\infty, \tag{2.3.9}$$

where c is a constant, $\alpha = \sqrt{\frac{\psi+\gamma^2/\sigma^2}{\sigma^2}}$, *and* $\beta = \frac{\gamma}{\sigma^2}$. *In addition, the right tail will be heavier if* $\gamma > 0$ *and the heavier right tail decays as*

$$f_{GH}(x; \lambda, \chi, \psi, \mu, \sigma, \gamma) \sim c x^{\lambda-1} e^{-\alpha x+\beta x} \text{ as } x \to \ +\infty \ \ if \gamma > 0, \tag{2.3.10}$$

and the heavier left tail decays as

$$f_{GH}(x; \lambda, \chi, \psi, \mu, \sigma, \gamma) \sim c|x|^{\lambda-1} e^{\alpha x+\beta x} \text{ as } x \to \ -\infty \ \ if \gamma < 0. \tag{2.3.11}$$

This theorem shows that the tail is the product of a power function and an exponential function. Generalized hyperbolic distributions can be regarded as semi-heavy tailed distributions.

Theorem 2.3.7 Tails of Skewed t Distribution. *If X is of skewed t distribution, then*

$$f(x; \nu, \mu, \sigma, \gamma) \sim c|x|^{-\nu/2-1} e^{-\alpha|x|+\beta x} \text{ as } x \to \pm\infty, \tag{2.3.12}$$

where c is a constant, $\alpha = \frac{|\gamma|}{\sigma^2}$, *and* $\beta = \frac{\gamma}{\sigma^2}$. *In addition, the heavier right tail decays as*

$$f(x; \nu, \mu, \sigma, \gamma) \sim c|x|^{-\nu/2-1} \text{ as } x \to \ +\infty \ \ if\ \gamma > 0, \tag{2.3.13}$$

and the heavier left tail decays as

$$f(x; \nu, \mu, \sigma, \gamma) \sim c|x|^{-\nu/2-1} \text{ as } x \to \ -\infty \ \ if\ \gamma < 0. \tag{2.3.14}$$

We can see that the tail of skewed t can decay as a power function and it is heavier than Student t distribution.

As we see before, there are six type of parameters in generalized hyperbolic distributions. We will show by some exhibits to see how these parameters play roles in the generalized hyperbolic distributions.

The Role of λ:

λ defines a sub-family of the generalized hyperbolic distributions. When $\lambda = 1$, it is hyperbolic distribution, and when $\lambda = -0.5$, it is normal inverse Gaussian distribution.

We set $\mu = 0, \gamma = 0, \psi = 1, \chi = 1$ and σ to be a constant so that the variance of the generalized hyperbolic distribution is 1 and mean is 0. In addition, we compare the density of standard Gaussian distribution with GH distributions.

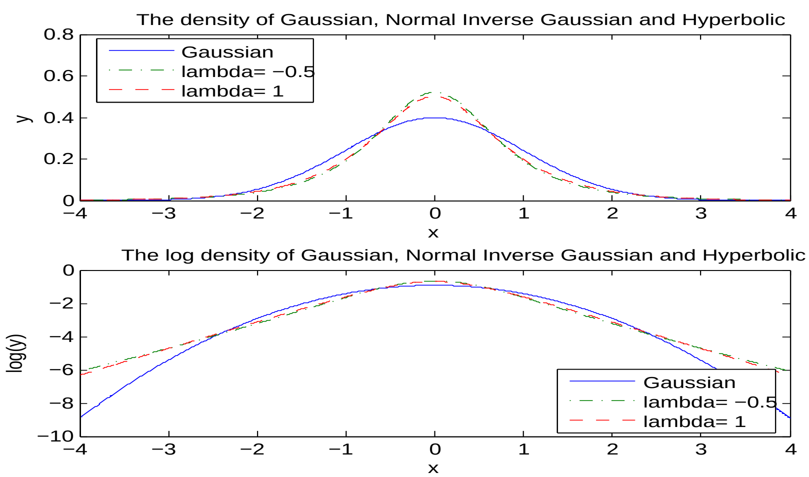

Figure 2.1: The density of GH and Gaussian

The log density of hyperbolic distribution is a hyperbola, while the log density of Gaussian distribution is a parabola. From Figure 2.1, We can see that hyperbolic, and normal inverse Gaussian obviously have heavier tails than Gaussian distribution. NIG has slightly heavier tails than Hyperbolic distribution(see McNeil et al.(2005)).

In the following exhibit, let us explore how the λ influences the tails and kurtosis. We let λ vary from -10 to 10, set $\mu = 0, \gamma = 0, \psi = 1, \chi = 1$ and σ to be a constant so that the variance of the generalized hyperbolic distribution is 1 and mean is 0.

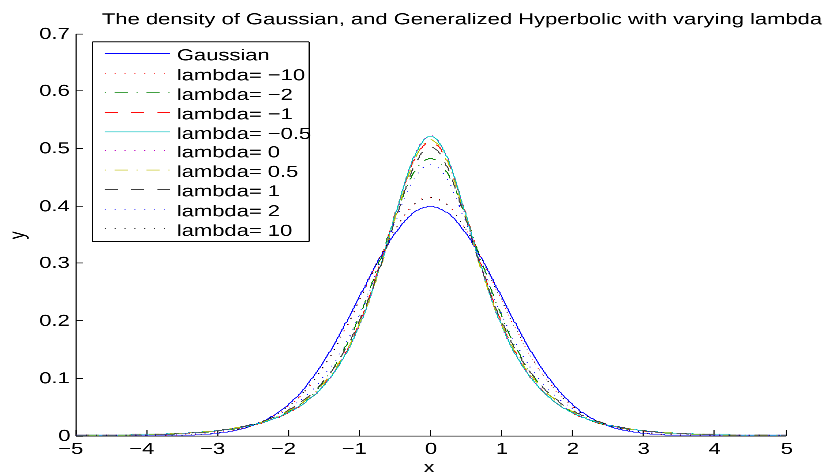

Figure 2.2: The density of GH and Gaussian

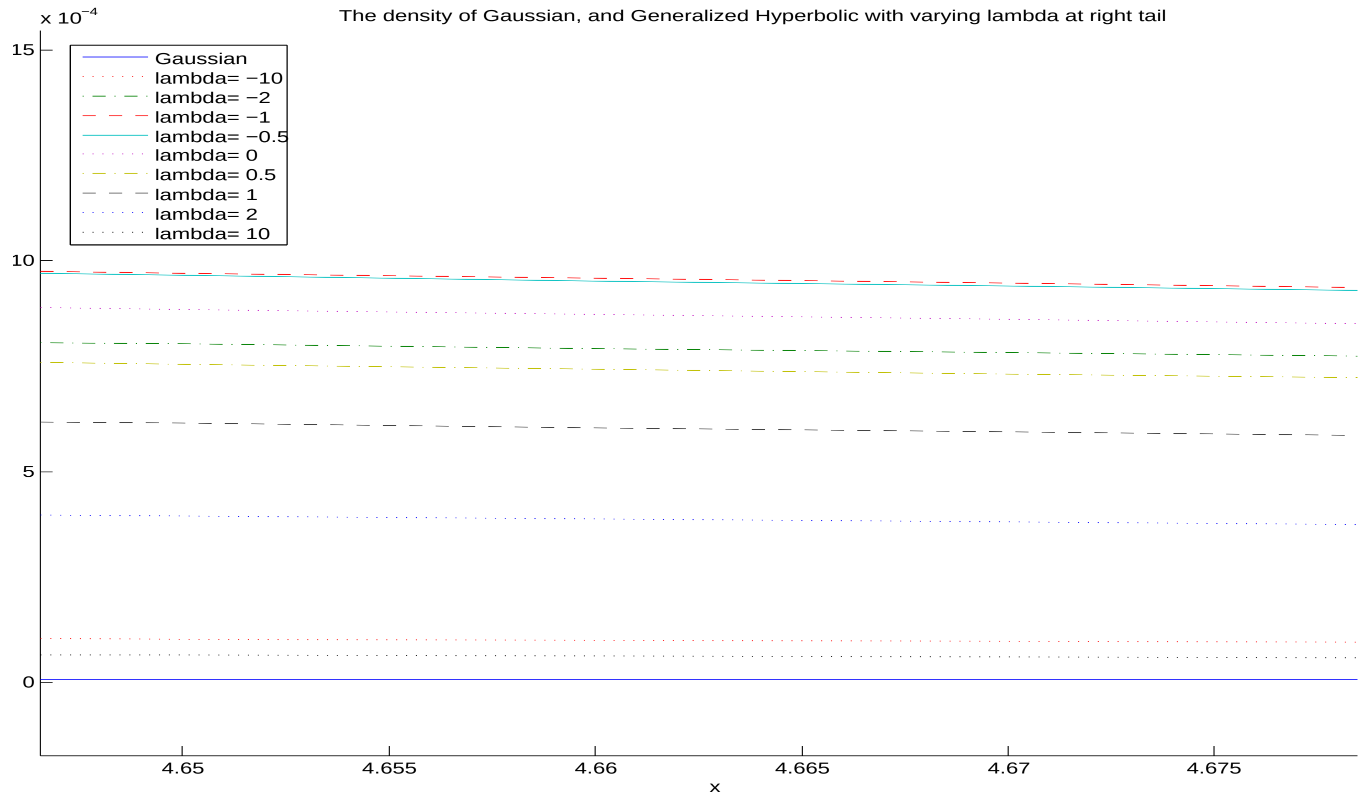

Figure 2.3: The density of GH and Gaussian at right tail

From Figure 2.2 and Figure 2.3, we can see that λ plays an important role in the generalized hyperbolic distributions. For all λ tested in this exhibit, all have heavier tails than Gaussian distribution and both NIG and $\lambda = -1$ cases have the heaviest tail among 9 tested generalized hyperbolic distributions. When $|\lambda|$ is small, the tails are heavy. When $|\lambda|$

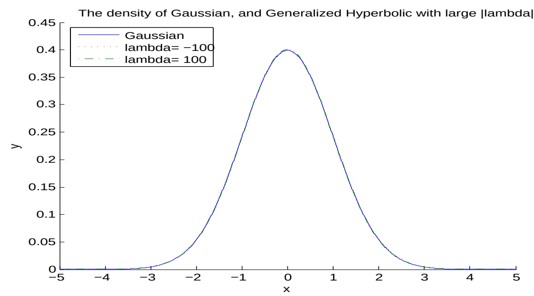

Figure 2.4: The density of GH with large $|\lambda|$and Gaussian

becomes much larger, the tails become thinner. Actually, when $|\lambda|$ is very large, symmetric generalized hyperbolic distributions look like the Gaussian distribution, which can be seen from figure 2.4.

The Role of χ:

We set $\mu = 0, \gamma = 0, \psi = 1, \lambda = 5$, let χ vary from 100 to 0.1 and σ to be a constant for the generalized hyperbolic distributions so that they have mean 0 and variance 1. We set $\mu = 0, \gamma = 0, \psi = 1, \lambda = 5$ and and σ to be a constant for the variance gamma distribution so that it has mean 0 and variance 1. From Figure 2.5, we can see that variance gamma distribution is the limiting distribution of generalized hyperbolic distributions when χ goes to 0, and variance gamma has the heaviest tail among those tested distributions.

The Role of ψ:

We set $\mu = 0, \gamma = 0, \chi = 5, \lambda = -2.5$, let ψ vary from 100 to 0.01 and σ to be a constant for the generalized hyperbolic distributions so that they have mean 0 and variance 1. We

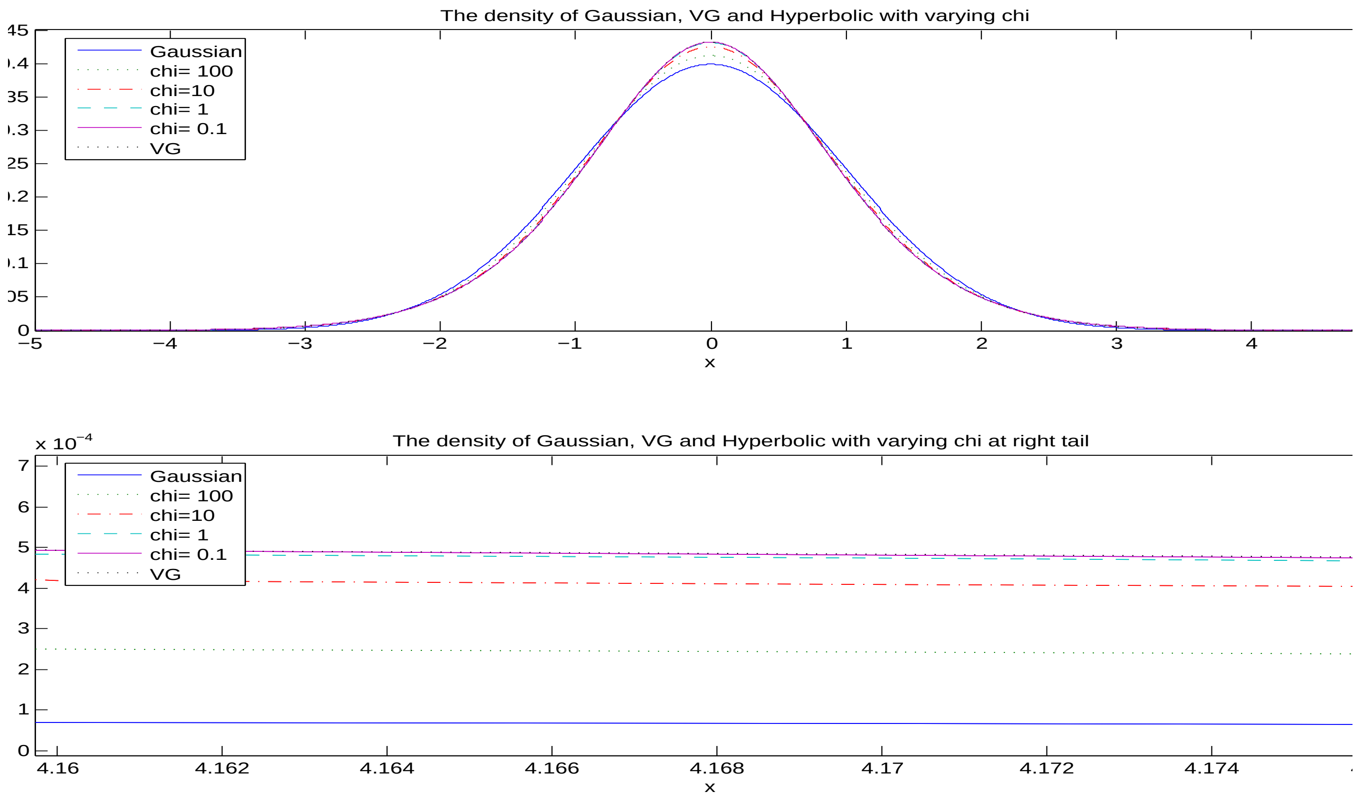

Figure 2.5: The density of Gaussian, GH with χ decreasing to 0 and variance gamma

set $\mu = 0, \gamma = 0, \nu = 5$ and σ to be a constant for the skewed t distribution so that it has mean 0 and variance 1.

From Figure 2.6, we can see that skewed t distribution is the limiting distribution of generalized hyperbolic distributions when ψ goes to 0, and $\lambda = -\nu/2$ and skewed t has the

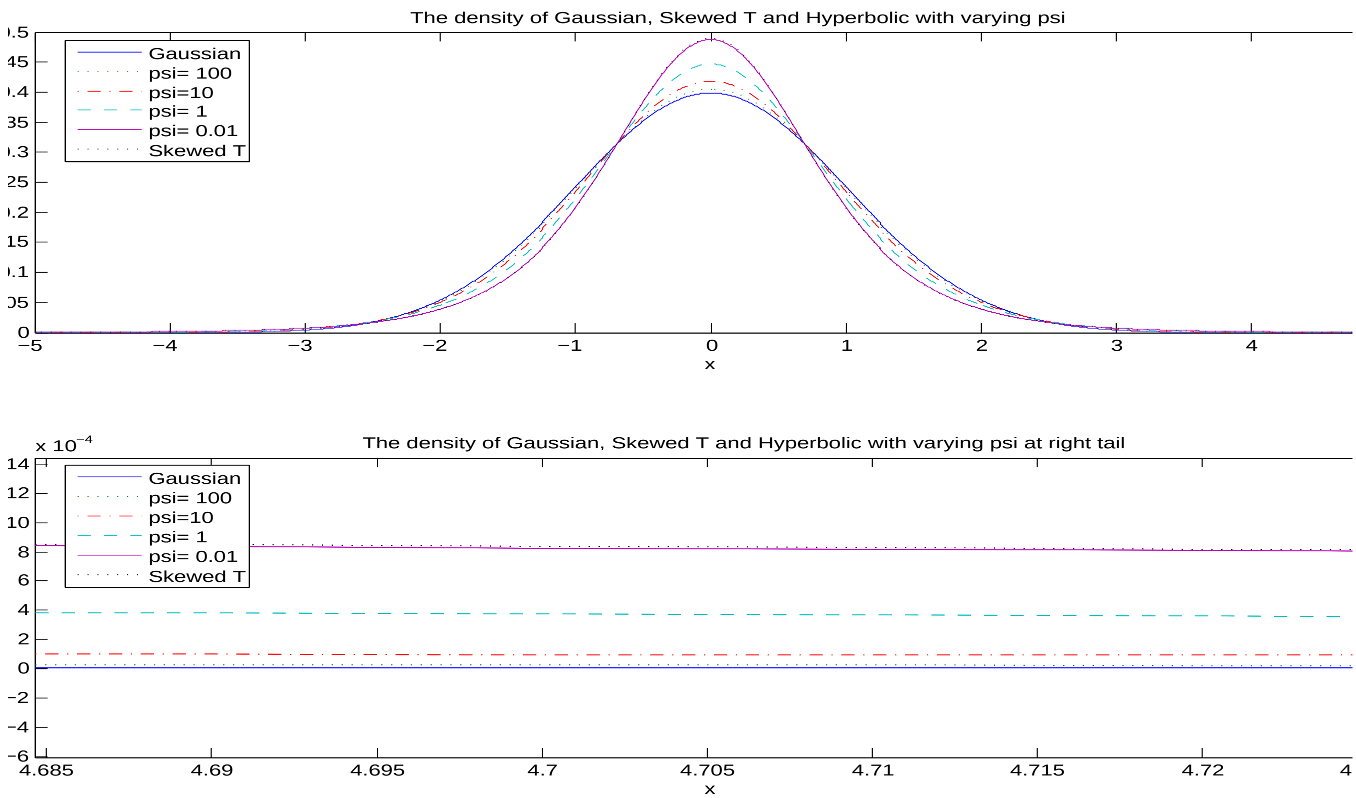

Figure 2.6: The density of Gaussian, GH with ψ decreasing to 0 and skewed t

heaviest tail among those tested distributions.

The Role of γ, μ, σ: μ is the location parameter, σ is the scale parameter and γ is skewness parameter. The location and scale parameters are common for all distributions. The skewness parameter is very useful in the modeling financial return data. It is believed that for a large part of the financial return data the left tail is slightly heavier than the right tail. From equation 2.3.9, we can see that when γ is negative, the left tail is slightly heavier than the right tail.

2.4 EM Algorithm for The Estimation of Generalized Hyperbolic Distributions

While univariate generalized hyperbolic models have been applied widely in the modeling of financial data, little work has been done about the multivariate generalized hyperbolic distributions. Maximizing the log-likelihood of multivariate generalized hyperbolic distributions is not very tractable. The conventional HYP program developed by Blæsid and Sörensen(1992) can not handle the calibration when the dimension is greater than three since there are too many parameters. Prause(1999) can only calibrate the symmetric generalized hyperbolic distributions at any dimension using HYP.

However, the mean-variance representation of generalized hyperbolic distributions has a great advantage. The EM algorithm can be applied to such a representation. The generalized hyperbolic random variable can be represented as a conditional normal distribution so that most of the parameters $(\Sigma, \boldsymbol{\mu}, \boldsymbol{\gamma})$ can be calibrated like a Gaussian distribution if other three type parameters (λ, χ, ψ) are already estimated or assumed to be some values.

We follow the EM algorithm framework of McNeil , Frey, and Embrechts(2005) for generalized hyperbolic distributions and provide our algorithm for generalized hyperbolic distributions with large $|\lambda|$ and call it ψ algorithm and χ algorithm. Importantly, we provide our special algorithms for the limiting or special cases: VG, skewed t, NIG, and Student t distributions.

Assume that we have *i.i.d.* data[‡] $\mathbf{x}_1, \cdots, \mathbf{x}_n$, where $\mathbf{x_i} \in \mathbb{R}^d$ and we want to fit these data by multivariate generalized hyperbolic distributions . The parameters are denoted by $\boldsymbol{\zeta} = (\lambda, \chi, \psi, \Sigma, \boldsymbol{\mu}, \boldsymbol{\gamma})$. The log-likelihood function that we want to maximize is

$$\log L(\boldsymbol{\zeta}; \mathbf{x}_1, \cdots, \mathbf{x}_n) = \sum_{i=1}^{n} \log f_{\mathbf{X_i}}(\mathbf{x}_i; \boldsymbol{\zeta}). \tag{2.4.1}$$

It is almost impossible to maximize the above objective function directly if the dimension is greater than three. The idea of the EM algorithm is to introduce the latent mixing variables

[‡]Usually, the financial data is not *i.i.d.*, we will solve this problem in Section 3.3 by using $GARCH$ filter.

$w_1, \cdots, w_n$ and suppose they were observable at the beginning and optimize the following quasi or augmented log-likelihood function,

$$\log \tilde{L}(\boldsymbol{\zeta}; \mathbf{x}_1, \cdots, \mathbf{x}_n, w_1, \cdots, w_n) = \sum_{i=1}^{n} \log f_{\mathbf{X_i}, W_i}(\mathbf{x_i}, w_i; \boldsymbol{\zeta}). \tag{2.4.2}$$

By the mean-variance mixture definition of generalized hyperbolic distributions, the log-likelihood function can be rewritten as

$$\begin{aligned} &\log \tilde{L}(\boldsymbol{\zeta}; \mathbf{x}_1, \cdots, \mathbf{x}_n, w_1, \cdots, w_n) = \\ &\quad \sum_{i=1}^{n} \log f_{\mathbf{X_i}|W_i}(\mathbf{x}_i|w_i; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) + \\ &\quad \sum_{i=1}^{n} \log h_{W_i}(w_i; \lambda, \chi, \psi) = \\ &\quad L_1(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}; \mathbf{x}_1, \cdots, \mathbf{x}_n | w_1, \cdots, w_n) + L_2(\lambda, \chi, \psi; w_1, \cdots, w_n), \end{aligned} \tag{2.4.3}$$

where $\mathbf{X}|W \sim N(\boldsymbol{\mu} + w\boldsymbol{\gamma}, w\Sigma)$ and $f_{\mathbf{X}|W}(x|w)$ is the density of conditional normal distribution, and $h(w)$ is the density function GIG distributed mixing random variable.

We can see from the above equation that the calibration of $\Sigma, \boldsymbol{\mu}, \boldsymbol{\gamma}$ and λ, χ, ψ can be separate by maximizing L_1, and L_2 respectively.

Following the same procedure in the proof of Theorem 2.2.2, the density of conditional normal distribution can be rewritten as,

$$f_{\mathbf{X}|W}(x|w) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} w^{\frac{d}{2}}} e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}} e^{\frac{-\rho}{2w}} e^{-\frac{w}{2}\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}},$$

where

$$\rho = (\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}).$$

Then, we can get the log-likelihood function L_1,

$$\begin{aligned} &L_1(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}; \mathbf{x}_1, \cdots, \mathbf{x}_n | w_1, \cdots, w_n) = \\ &\quad -\frac{n}{2} \log |\Sigma| - \frac{d}{2} \sum_{i=1}^{n} \log w_i + \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma} \\ &\quad -\frac{1}{2} \sum_{i=1}^{n} \frac{1}{w_i}\rho_i - \frac{1}{2}\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma} \sum_{i=1}^{n} w_i. \end{aligned} \tag{2.4.4}$$

From equation 2.1.10, we can get the log-likelihood function L_2,

$$L_2(\lambda, \chi, \psi; w_1, \cdots, w_n) = \\ (\lambda - 1)\sum_{i=1}^{n} \log w_i - \frac{\chi}{2}\sum_{i=1}^{n} w_i^{-1} - \frac{\psi}{2}\sum_{i=1}^{n} w_i - \frac{n\lambda}{2}\log\chi \\ + \frac{n\lambda}{2}\log\psi - n\log\left(2K_\lambda(\sqrt{\chi\psi})\right). \tag{2.4.5}$$

Estimation of $\Sigma, \boldsymbol{\mu}, \boldsymbol{\gamma}$ is obtained by maximizing L_1. Suppose that the latent mixing variables $w_1, \cdots, w_n$ are observable. Following the standard routine of optimization, we take the partial derivative of L_1 with respect to Σ, $\boldsymbol{\mu}$, and $\boldsymbol{\gamma}$, and set

$$\frac{\partial L_1}{\partial \boldsymbol{\mu}} = \mathbf{0} \\ \frac{\partial L_1}{\partial \boldsymbol{\gamma}} = \mathbf{0} \\ \frac{\partial L_1}{\partial \Sigma} = \mathbf{0}. \tag{2.4.6}$$

From the above equation array, we can get the following estimations[§],

$$\boldsymbol{\gamma} = \frac{n^{-1}\sum_{i=1}^{n} w_i^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{x}_i)}{n^{-2}\left(\sum_{i=1}^{n} w_i\right)\left(\sum_{i=1}^{n} w_i^{-1}\right) - 1} \tag{2.4.7}$$

$$\boldsymbol{\mu} = \frac{n^{-1}\sum_{i=1}^{n} w_i^{-1}\boldsymbol{x}_i - \boldsymbol{\gamma}}{n^{-1}\sum_{i=1}^{n} w_i^{-1}} \tag{2.4.8}$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} w_i^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})' - \frac{1}{n}\sum_{i=1}^{n} w_i\boldsymbol{\gamma}\boldsymbol{\gamma}'. \tag{2.4.9}$$

Estimation of λ, χ, ψ is obtained by maximizing L_2. In general, we assume λ to be a constant. Nobody reports that one can calibrate λ for multivariate generalized hyperbolic distributions until now. To maximize L_2, we take the partial derivative with respect to χ and ψ and solve the following equation array,

$$\frac{\partial L_2}{\partial \chi} = 0 \\ \frac{\partial L_2}{\partial \psi} = 0. \tag{2.4.10}$$

[§]To solve this equation array, some knowledge of multivariate statistics algebra is required. Most standard multivariate statistics books have this in the introduction chapter or appendix.

Solving the above equation array leads us to solve $\theta = \sqrt{\chi\psi}$ from the following equation first,

$$n^{-2}\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j^{-1} K_\lambda^2(\theta)\theta + 2\lambda K_{\lambda+1}(\theta)K_\lambda(\theta) - \theta K_\lambda^2(\theta) = 0^{\P}. \quad (2.4.11)$$

We find θ by zero-finder routine in Matlab, fzero. Once θ is solved, we can get parameters (χ, ψ),

$$\chi = \frac{n^{-1}\theta\sum_{i=1}^{n} w_i K_\lambda(\theta)}{K_{\lambda+1}(\theta)}, \quad (2.4.12)$$

$$\psi = \frac{\theta^2}{\chi}. \quad (2.4.13)$$

Especially, when $\lambda = -0.5$, we have the normal inverse Gaussian distribution, and we are able to get θ explicitly since $K_{-\lambda}(x) = K_\lambda(x)$ for any λ,

$$\theta = \frac{2\lambda}{1 - n^{-2}\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j^{-1}}. \quad (2.4.14)$$

Therefore, for NIG, we have a comparatively fast algorithm. Protassov(2004) successfully fit NIG to five dimensional currency data.

When $\chi = 0$, and $\lambda > 0$, it is variance gamma distribution and we are able to get λ by setting $\frac{\partial L_2}{\partial \lambda} = 0$ from the following equation,

$$\log(\lambda) - \log(n^{-1}\sum_{i=1}^{n} w_i) + n^{-1}\sum_{i=1}^{n}\log(w_i) - \phi(\lambda) = 0, \quad (2.4.15)$$

where $\phi(\lambda)$ is the di-gamma function. We can get

$$\psi = \frac{2\lambda}{n^{-1}\sum_{i=1}^{n} w_i}. \quad (2.4.16)$$

When $\psi = 0$, it is an unknown distribution, however, if we continue to set $\lambda = -\nu/2$, and $\chi = \nu$, we can get skew t distribution with degree of freedom ν. Following standard routine mentioned above, the only type parameter, ν, can be solved from the following equation,

$$-\psi(\frac{\nu}{2}) + \log(\nu/2) + 1 - n^{-1}\sum_{i=1}^{n} w_i^{-1} - n^{-1}\sum_{i=1}^{n}\log(w_i) = 0. \quad (2.4.17)$$

However, the latent mixing variables $w_1, \cdots, w_n$ are not observable. An iteration procedure consisting of E-step and M-step is needed. The E-step is called the estimation

¶This equation is similar to equation 2.4.36 and we prove equation 2.4.36 in the appendix.

step. In this step, the conditional expectation of the augmented log-likelihood function given current parameter estimates and sample data is calculated. Suppose that we are at step k, we need to calculate the following conditional expectation and get a new objective function to be maximized,

$$Q(\boldsymbol{\zeta};\boldsymbol{\zeta}^{[k]}) = E\left(\log \tilde{L}(\boldsymbol{\zeta};\mathbf{x}_1,\cdots,\mathbf{x}_n,W_1,\cdots,W_n)|\mathbf{x}_1,\cdots,\mathbf{x}_n;\boldsymbol{\zeta}^{[k]}\right).$$

In the M-step, we maximize the above new objection function to get updated estimates $\boldsymbol{\zeta}^{[k+1]}$. From equation 2.4.4 and 2.4.5, we can see that it is equivalent to updating all the w_i, w_i^{-1}, and $\log(w_i)$ in the augmented log-likelihood function by their conditional estimates $E(W_i|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\boldsymbol{k}]})$, $E(W_i^{-1}|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\boldsymbol{k}]})$, and $E(\log(W_i)|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\boldsymbol{k}]})$. In this way, $Q(\boldsymbol{\zeta};\boldsymbol{\zeta}^{[k]})$ is expressed by observations and known conditional expectations so that it can be maximized as we have just done. To calculate those conditional expectations, we need the following conditional density function,

$$f_{W|\boldsymbol{X}}(w|\boldsymbol{x};\boldsymbol{\zeta}) = \frac{f(\boldsymbol{x}|w;\boldsymbol{\zeta})h(w;\boldsymbol{\zeta})}{f(\boldsymbol{x};\boldsymbol{\zeta})}$$

By some algebra work, we can get,

$$W_i|\boldsymbol{X}_i \;\sim\; N^{-}(\lambda-\frac{d}{2},\rho_i+\chi,\psi+\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}). \tag{2.4.18}$$

In the following, we calculate $E(W_i|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\boldsymbol{k}]})$, $E(W_i^{-1}|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\boldsymbol{k}]})$, and $E(\log(W_i)|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\boldsymbol{k}]})$ for generalized hyperbolic distributions and the limiting distributions such as variance gamma, skewed t and Student t etc.

For convenience, we use the standard notation of Liu and Rubin(1995), Protassov(2004) and McNeil, Frey, and Embrechts(2005),

$$\delta_i^{[\cdot]} = E\left(W_i^{-1}|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\cdot]}\right),\ \eta_i^{[\cdot]} = E\left(W_i|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\cdot]}\right),\ \xi_i^{[\cdot]} = E\left(\log(W_i)|\boldsymbol{x}_i;\boldsymbol{\zeta}^{[\cdot]}\right), \tag{2.4.19}$$

and

$$\bar{\delta} = \frac{1}{n}\sum_1^n \delta_i,\ \bar{\eta} = \frac{1}{n}\sum_1^n \eta_i,\ \bar{\xi} = \frac{1}{n}\sum_1^n \xi_i. \tag{2.4.20}$$

For the generalized hyperbolic distributions, by using equation 2.1.11 and 2.1.12 we get,

$$\delta_i^{[k]} = \left(\frac{\rho_i^{[k]}+\chi^{[k]}}{\psi^{[k]}+\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]}}\right)^{-\frac{1}{2}} \frac{K_{\lambda-\frac{d}{2}-1}\left(\sqrt{(\rho_i^{[k]}+\chi^{[k]})(\psi^{[k]}+\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\rho_i^{[k]}+\chi^{[k]})(\psi^{[k]}+\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)} \tag{2.4.21}$$

$$\eta_i^{[k]} = \left(\frac{\rho_i^{[k]} + \chi^{[k]}}{\psi^{[k]} + \boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]}}\right)^{\frac{1}{2}} \frac{K_{\lambda-\frac{d}{2}+1}\left(\sqrt{(\rho_i^{[k]} + \chi^{[k]})(\psi^{[k]} + \boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\rho_i^{[k]} + \chi^{[k]})(\psi^{[k]} + \boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)} \quad (2.4.22)$$

$$\xi_i^{[k]} = \frac{1}{2}\log\left(\frac{\rho_i^{[k]} + \chi^{[k]}}{\psi^{[k]} + \boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]}}\right) + \frac{\frac{\partial K_{\lambda-\frac{d}{2}+\alpha}\left(\sqrt{(\rho_i^{[k]}+\chi^{[k]})(\psi^{[k]}+\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}{\partial\alpha}|_{\alpha=0}}{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\rho_i^{[k]} + \chi^{[k]})(\psi^{[k]} + \boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}. \quad (2.4.23)$$

We use numerical method to calculate above derivatives at zero. In general, we do not need $\xi_i^{[k]}$ when λ is fixed. We do need it in the calibration of skewed t distribution mentioned later.

For the variance gamma distribution, we have

$$W_i|\boldsymbol{X}_i \;\sim\; N^{-}(\lambda - \frac{d}{2}, \rho_i, \psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}), \quad (2.4.24)$$

and we just need to set $\chi = 0$ in all the above three equations.

For the multivariate skewed t distributions, we have

$$W_i|\boldsymbol{X}_i \;\sim\; N^{-}(-\frac{d+\nu}{2}, \rho_i + \nu, \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}), \quad (2.4.25)$$

and

$$\delta_i^{[k]} = \left(\frac{\rho_i^{[k]} + \nu^{[k]}}{\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]}}\right)^{-\frac{1}{2}} \frac{K_{\frac{\nu+d+2}{2}}\left(\sqrt{(\rho_i^{[k]} + \nu^{[k]})(\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}{K_{\frac{\nu+d}{2}}\left(\sqrt{(\rho_i^{[k]} + \nu^{[k]})(\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)} \quad (2.4.26)$$

$$\eta_i^{[k]} = \left(\frac{\rho_i^{[k]} + \nu^{[k]}}{\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]}}\right)^{\frac{1}{2}} \frac{K_{\frac{\nu+d-2}{2}}\left(\sqrt{(\rho_i^{[k]} + \nu^{[k]})(\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}{K_{\frac{\nu+d}{2}}\left(\sqrt{(\rho_i^{[k]} + \nu^{[k]})(\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)} \quad (2.4.27)$$

$$\xi_i^{[k]} = \frac{1}{2}\log\left(\frac{\rho_i^{[k]} + \nu^{[k]}}{\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]}}\right) + \frac{\frac{\partial K_{-\frac{\nu+d}{2}+\alpha}\left(\sqrt{(\rho_i^{[k]}+\nu^{[k]})(\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}{\partial\alpha}|_{\alpha=0}}{K_{\frac{\nu+d}{2}}\left(\sqrt{(\rho_i^{[k]} + \nu^{[k]})(\boldsymbol{\gamma}^{[k]\prime}\Sigma^{[k]-1}\boldsymbol{\gamma}^{[k]})}\right)}. \quad (2.4.28)$$

Especially, for the popular multivariate Student t distribution, i.e., when $\boldsymbol{\gamma} = 0$, the conditional distributions of latent mixing variables are in the following simpler form,

$$W_i|\boldsymbol{X}_i \ \sim \ InverseGamma(\frac{d+\nu}{2}, \frac{\rho_i+\nu}{2}). \tag{2.4.29}$$

Thus we have greatly simplified formulas for those conditional expectation estimates,

$$\delta_i^{[k]} = \frac{\nu^{[k]}+d}{\rho_i^{[k]}+\nu^{[k]}} \tag{2.4.30}$$

$$\eta_i^{[k]} = \frac{\rho_i^{[k]}+\nu^{[k]}}{\nu^{[k]}+d-2} \tag{2.4.31}$$

$$\xi_i^{[k]} = \log(\frac{\rho_i^{[k]}+\nu^{[k]}}{2}) - \psi(\frac{d+\nu^{[k]}}{2}). \tag{2.4.32}$$

In the M-step, we just need to replace the latent variables w_i^{-1} by $\delta_i^{[k]}$, w_i by $\eta_i^{[k]}$, $\log(w_i)$ by $\xi_i^{[k]}$ in the maximization.

From the maximization of conditional expectation of L_1, we can get the following k-step estimations,

$$\boldsymbol{\gamma}^{[k+1]} = \frac{n^{-1}\sum_{i=1}^n \delta_i^{[k]}(\bar{\boldsymbol{x}} - \boldsymbol{x}_i)}{\bar{\delta}^{[k]}\bar{\eta}^{[k]} - 1} \tag{2.4.33}$$

$$\boldsymbol{\mu}^{[k+1]} = \frac{n^{-1}\sum_{i=1}^n \delta_i^{[k]}\boldsymbol{x}_i - \boldsymbol{\gamma}^{[k+1]}}{\bar{\delta}^{[k]}} \tag{2.4.34}$$

$$\Sigma^{[k+1]} = \frac{1}{n}\sum_{i=1}^n \delta_i^{[k]}(\boldsymbol{x}_i - \boldsymbol{\mu}^{[k+1]})(\boldsymbol{x}_i - \boldsymbol{\mu}^{[k+1]})' - \bar{\eta}^{[k]}\boldsymbol{\gamma}^{[k+1]}\boldsymbol{\gamma}^{[k+1]\prime}. \tag{2.4.35}$$

From the maximization of conditional expectation of L_2, we can get the estimation of $\theta^{[k+1]}$ from the following equation first,

$$\bar{\eta}^{[k]}\bar{\delta}^{[k]}K_\lambda^2(\theta)\theta + 2\lambda K_{\lambda+1}(\theta)K_\lambda(\theta) - \theta K_\lambda^2(\theta) = 0^{\|}. \tag{2.4.36}$$

We can get other parameters by

$$\chi^{[k+1]} = \frac{\theta^{[k+1]}\bar{\eta}^{[k]}K_\lambda(\theta^{[k+1]})}{K_{\lambda+1}(\theta^{[k+1]})}, \tag{2.4.37}$$

$$\psi^{[k+1]} = \frac{\theta^{[k+1]^2}}{\chi^{[k+1]}}. \tag{2.4.38}$$

‖Proof can be found in the appendix.

There is an identification problem for generalized hyperbolic distributions. McNeil et al.(2005) set the determinant of Σ to be c, the determinant of sample covariance matrix to solve this problem by using equation 2.2.12 and set

$$\Sigma^{[k+1]} := \frac{c^{1/d}\Sigma^{[k+1]}}{\left|\Sigma^{[k+1]}\right|^{1/d}} \tag{2.4.39}$$

From our calibration experience, there is a problem here. When $|\lambda|$ is large, there may be no zeros in equation 2.4.36 so that the program will crash.

We find out that when λ is greater than a certain number(say -1), we set χ to be a constant and solve $\theta^{[k+1]}$ from following equation,

$$\theta\bar{\eta}^{[k]}K_\lambda(\theta) - K_{\lambda+1}(\theta)\chi = 0 \tag{2.4.40}$$

and solve $\psi^{[k+1]}$ from

$$\psi^{[k+1]} = \frac{\theta^{[k+1]^2}}{\chi}, \tag{2.4.41}$$

and that when λ is smaller than a certain number(say 1), we set ψ to be a constant and solve $\theta^{[k+1]}$ from following equation,

$$\theta\bar{\delta}^{[k]}K_\lambda(\theta) - K_{\lambda-1}(\theta)\psi = 0^{**} \tag{2.4.42}$$

and solve $\chi^{[k+1]}$ from

$$\chi^{[k+1]} = \frac{\theta^{[k+1]^2}}{\psi}. \tag{2.4.43}$$

We set χ to be some constant and call this the χ algorithm when λ is larger than some number(say -1) and set ψ to be some constant when λ is less than some number(say 1) and call this the ψ algorithm. From our experience, different choices of constant χ or ψ lead to different calibration speeds.

It is noted that Protassov(2004) set χ to be 1 to simplify the derivations although he did not mention this. In addition, his framework is in the usual definition of GH. From our experience, his algorithm is not stable when λ is small. His algorithm is a special case of our algorithm when λ is greater than a certain number. Our algorithm can calibrate all the sub-family defined by $\lambda \in \mathbb{R}$. To our knowledge, we are the first who can calibrate the generalized hyperbolic distributions when $|\lambda|$ is large[††].

[**] Proof of this equation and equation 2.4.40 can be found in the appendix.

[††] When $|\lambda|$ is greater than a very large number(say 100), the Bessel function is non tractable. Usually, we set $|\lambda| < 10$.

For normal inverse Gaussian distribution, we can get $\theta^{[k+1]}$ explicitly,

$$\theta^{[k+1]} = \frac{2\lambda}{1 - \bar{\eta}^{[k]}\bar{\delta}^{[k]}}. \tag{2.4.44}$$

For variance gamma distribution, we can get $\lambda^{[k+1]}$ from following equation,

$$\log(\lambda) - \log(\bar{\eta}^{[k]}) + \bar{\xi}^{[k]} - \phi(\lambda) = 0, \tag{2.4.45}$$

and

$$\psi^{[k+1]} = \frac{2\lambda}{\bar{\eta}^{[k]}}. \tag{2.4.46}$$

For skewed t distribution, degree of freedom ν can be solved from the following equation,

$$-\psi(\frac{\nu}{2}) + \log(\nu/2) + 1 - \bar{\xi}^{[k]} - \bar{\delta}^{[k]} = 0. \tag{2.4.47}$$

The calibration of multivariate Student t, i.e., when $\boldsymbol{\gamma} = 0$, can also use above procedures. In addition, we can maximize the original log-likelihood function given current estimates to get a fast algorithm. This is called $ECME$ algorithm. In this case, we can solve the following equation for ν to get $\nu^{[k+1]}$,

$$\psi(\frac{\nu+d}{2}) - \psi(\frac{\nu}{2}) + \log(\frac{\nu}{\nu+d}) + 1 - \frac{1}{n}\sum_{i=1}^{n}\left(\log\delta_i^{[k]} - \delta_i^{[k]}\right) = 0. \tag{2.4.48}$$

More details of $ECME$ can be found in Liu and Rubin(1995).

After we update all the parameters, one iteration of standard EM algorithm is completed. Some argue that we should recalculate δ, η and ξ, and call them $\delta_i^{[k,1]}$, $\eta_i^{[k,1]}$, and $\xi_i^{[k,1]}$ after the k-step estimation of $\Sigma^{[k+1]}$, $\boldsymbol{\gamma}^{[k+1]}$, and $\boldsymbol{\gamma}^{[k+1]}$ to estimate $\chi^{[k+1]}$, $\psi^{[k+1]}$ and $\lambda^{[k+1]}$(in some cases), and it is called $MCECM$ algorithm. We do not use this technique for variance gamma distribution and skewed t distribution since the calibration of those two distributions will need to calculate the extra ξ, which may lower the calibration speed. McNeil, Frey, and Embrechts(2005) argued that we might be able to maximize the original log-likelihood for the generalized hyperbolic distributions as in Student t case in the estimation of $\chi^{[k+1]}$, $\psi^{[k+1]}$ and $\lambda^{[k+1]}$, and it is called $ECME$ algorithm. From our calculation, $ECME$ and EM algorithm are equivalent for generalized hyperbolic distributions, and both algorithms will lead to equation 2.4.36 to solve θ.

The iteration will stop if the relative increment of log-likelihood is trivial.

2.5 Empirical Experiments

2.5.1 Fitting Financial Market Data

In the maximization of L_2, McNeil et al(2005) set the determinant of dispersion matrix to be the determinant of sample covariance matrix and argued that that this usually gives a stable performance for the algorithms. Protassov(2004) set χ to be 1 though he did not mention this and his algorithm is in the usual parametrization.

We compare the performance of those four algorithms(McNeil et al., Protassov, χ and ψ algorithm) for one dimensional and five dimensional hyperbolic distribution [‡‡]. We also compare the performance of the limiting or special cases, VG, skewed t and NIG distributions. Table 2.1 shows the time spent in calibrating the one dimensional or five dimensional distribution and the corresponding log likelihood. This table shows that McNeil et al (2005) algorithm is quite stable and fast for hyperbolic distribution since their L_2 is optimized over two variables $(\chi, \psi)$and rescale all the parameters at the end of each EM iteration step. Our algorithm and Protassov's algorithm use the same idea and L_2 is optimized over one variable, though Protassov set χ to be 1 to simply the derivation of the algorithm and his algorithm is in the usual parametrization. If we choose suitable parameters, our algorithm may outperform McNeil et al(2005) and Protassov(2004).

Table 2.1: Calculation time and log likelihood for generalized hyperbolic distributions

Model	1d likelihood	time(seconds)	5d likelihood	time
Hyperbolic(McNeil et al.)	-1043.49	65	-4891.46	17
Hyperbolic(Protassov)	-1043.84	690	-4891.46	340
Hyperbolic(χ)	-1043.50	18	-4891.15	250
Hyperbolic(ψ)	-1043.52	18	-4891.46	95
VG	-1044.97	130	-4901.74	28
NIG	-1042.76	140	-4884.15	28
Skewed t	-1039.70	118	-4873.91	27

[‡‡]We use the data from chapter 4, i.e., most recent five dimensional 750 filtered *i.i.d.* returns data. For the one dimensional calibration, we use the first column of the data. These four algorithms use the same initial parameters and the same termination condition(For one dimension, the absolute relative increment of log likelihood is less than than e^{-6}, while for five dimension, e^{-8}). We set χ to be 5 for 1d and 1.5 for 5d. We set ψ to be 5 for 1d and 3 for 5d. We use a laptop with centrino 1.3GHZ CPU and 1GB PC2700 memory. Software is Matlab R14.

However, McNeil et al(2005)'s algorithm will crash when $|\lambda|$ is greater than some number(say 2.5) since there is no zero in equation 2.4.36. Protassov(2004)'s algorithm will crash too when λ is less than some number(say -4). Our algorithm is stable. We can calibrate any generalized hyperbolic distribution. In figure 2.7, we plot the log likelihood of calibrated one dimensional generalized hyperbolic distributions versus λ. From this figure, we can see that the generalized hyperbolic distribution with $\lambda = -5$ or -6 has highest log likelihood. This method can be regarded as a rough method to calibrate λ. Both Protassov(2004) and McNeil et al.(2005) could not calibrate the generalized hyperbolic distribution when $\lambda = -5$ or -6.

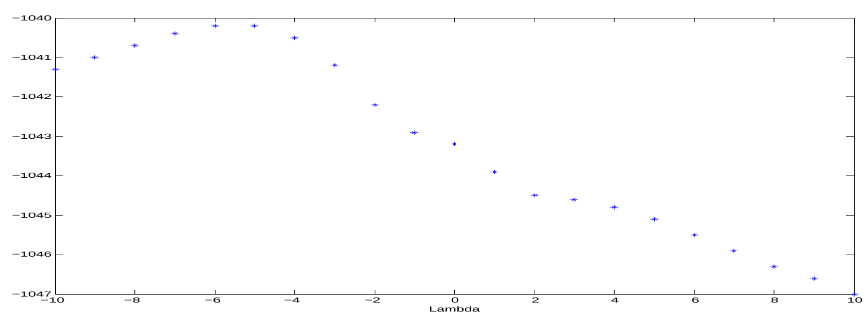

Figure 2.7: Log likelihood of generalized hyperbolic distributions versus λ

A hybrid method may consider to use χ or ψ algorithm at the beginning phase of calibration and use McNeil et al.(2005)'s algorithm later to get fast algorithm.

Furthermore, from table 2.1 and figure 2.7, we can see that skewed t distribution has fewest number of parameters and largest log likelihood comparing with GH when we loop λ from -10 to 10 with step size 1 and the calibration is fast so that skewed t is very promising in the applications.

2.5.2 Fitting Simulated Skewed t

In this section, we test the validity of our calibration algorithm for the case of the skewed t distribution, which is relatively easy to simulate. We generate simulated data from a known choice of skewed t parameters, and then check to see that our calibration algorithm correctly recovers those parameters.

We simulate 10,000 samples from two dimensional skewed t with parameters $(\mu_1, \mu_2)^T = (0, 0)^T$, $(\gamma_1, \gamma_2)^T = (-0.2, 0.2)^T$, $\nu = 6$ and $\Sigma = [1, 0.5; 0.5, 1]$ and fit the simulated 10,000 samples by skewed t distribution. We try 10 simulations and take the average of those 10 estimations. The calibrated parameters $(\mu_1, \mu_2)^T = (-0.007, 0.003)^T$, $(\gamma_1, \gamma_2)^T = (-0.192, 0.199)^T$, $\nu = 5.993$ and $\Sigma = [0.999, 0.503; 0.503, 1.012]$.

The results of this experiment support the validity of our EM algorithm for skewed t distributions, and lend confidence to the general case.

2.6 Conclusion

We follow McNeil, Frey and Embrechts(2005)'s parametrization to introduce generalized hyperbolic distributions from generalized inverse Gaussian distributions. The linear transformation of generalized hyperbolic distributions remains in the same sub-family of generalized hyperbolic distributions. Generalized hyperbolic distributions have semi-heavy tails so that they may be good candidates for risk management. We follow Liu and Rubin(1995), McNeil, Frey and Embrechts(2005) and Protassov(2004)'s EM algorithm and provide our EM algorithm to calibrate multivariate generalized hyperbolic distributions and all the well known limiting distributions. We find that the algorithm of Protassov(2004) under usual parametrization is a special case of our algorithm. From our experience, the algorithms in McNeil, Frey, and Embrechts(2005)and Protassov(2004) sometimes are not stable , while our algorithm is stable. We provide fast algorithms for NIG and skewed t under this new parametrization. We usually fix λ to calibrate the generalized hyperbolic distributions. We can also find λ by looping. For variance gamma distribution, we can calibrate λ. In addition, we find that skewed t distribution has fewest number of parameters and larger log likelihood with fast calibration speed so that skewed t distribution is very promising in the applications.

CHAPTER 3

A $GARCH - GH$ model for VaR Risk Management

Value at Risk (VaR) based on the normal distribution has been considered as the standard risk measure since J.P. Morgan launched RiskMetrics in 1994. In 1995, the Basel Committee on Banking Supervision suggested using the 10 day VaR at the 99% level. If the underlying loss distribution is strictly increasing, then VaR at confidence level α is defined as

$$VaR_\alpha = F^{-1}(\alpha), \tag{3.0.1}$$

where F^{-1} is the inverse of the cumulative distribution function of the loss random variable. We can see that VaR_α depends only on the choice of underlying distribution.

Yet, financial return series is heavy tailed and leptokurtic; large losses occur far away from the VaR based on normal distribution. At the 95% level, VaR based on a normal distribution is usually good enough. However, at the 99% level, VaR based on a normal distribution usually underestimates the true risk. Risk managers can not neglect this problem, and generalized hyperbolic distributions(GH) has been considered as an alternative. The tails of GH decay as $|x|^{\lambda-1}e^{-\alpha|x|+\beta x}$. Generalized hyperbolic distributions have at least semi-heavy tails and both tails decay as the product of a power function and an exponential function. For the limiting distribution known as the skewed t distribution, the tail can even decay as a power function $x^{-\nu/2-1}$ and it is heavier than the symmetric Student t, whose tails decay as $x^{-\nu-1}$ (The Gaussian distribution's tails decay as $e^{-x^2/2}$ and the tails are very thin compared to generalized hyperbolic distributions.).

Two special cases of generalized hyperbolic distributions: normal inverse Gaussian (NIG) and hyperbolic distributions are widely used in the modeling of univariate financial data for their better fit. Barndorff-Nielsen (1997) argued that the NIG is slightly superior to the hyperbolic as a univariate model for return data. Two limiting cases: skewed t and variance gamma distribution(VG) are rarely known in the modeling of financial data. In fact, both limiting cases have heavier tails than the Gaussian distribution, and they have fast algorithms calibrating all the parameters. Barbachan, Farias and Ornelas (2003) and Fajardo and Farias (2003) used univariate generalized hyperbolic distributions to model the Brazilian data to get more accurate VaR measurements. Aas and Hobæk Haff(2005a) used skewed t to model VaR (All previous work uses the usual parametrization.). We use the new parametrization of generalized hyperbolic distributions to fit the univariate return series and calculate the corresponding VaR_α in this chapter.

In the last chapter we developed the EM algorithms for the multivariate generalized hyperbolic distributions and their limiting cases. If we set the dimension to be one, then the algorithm becomes the one dimensional EM algorithm. In this chapter we model the negative log return of the stock index S&P500 and Dow Jones Industrial Average, respectively, by using generalized hyperbolic distributions.

The maximization of likelihood assumes that the return series is $i.i.d.$. However, typical financial return series usually is not independently and identically distributed ($i.i.d.$), so before modeling returns with a particular choice of distribution, we need to filter the data. We use a $GARCH(1,1)$ filter with Student t innovations to filter our data.

After we get the $i.i.d.$ return series, we calibrate NIG, VG, skewed t and hyperbolic distributions chosen from the general GH family and then implement a VaR backtesting procedure suggested by McNeil(1999). We show that use of GH distributions leads to better VaR forecasts than Gaussian.

This chapter is organized as following. We give a precise definition of VaR in Section 3.1. We plot the autocorrelation function(ACF) of return series in 3.2 and we will see that the return series is not $i.i.d.$. We solve this problem by introducing a $GARCH$ filter with student t innovations in Section 3.3 to get filtered return series. The filtered return series is approximately $i.i.d.$. In Section 3.4, we use QQ-plots to show that the normal distribution is too thin-tailed. We use heavy tailed generalized hyperbolic distributions in Section 3.5 to fit the univariate filtered S&P500 and Dow index, respectively. We backtest and compare VaR forecasts based on generalized hyperbolic distributions and normal distribution in Section 3.6.

3.1 Value at Risk(VaR)

Definition 3.1.1 Value at Risk(VaR). *VaR at confidence level $\alpha \in (0,1)$ for loss L of a security or a portfolio is defined to be*

$$VaR_\alpha = inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}, \tag{3.1.1}$$

where F is the distribution function of loss L.

If the loss distribution function F is strictly increasing, then $VaR_\alpha = F^{-1}(\alpha)$. In practice, the confidence level ranges from 95% through 99.5%, though Basel committee recommends 99%. Sometimes, $VaR_\alpha^{mean} := VaR_\alpha - \mu$, where μ is the mean of loss, is used instead of ordinary VaR.

If L is of normal distribution $N(\mu, \sigma^2)$, then

$$VaR_\alpha = \mu + \sigma\Phi^{-1}(\alpha), \tag{3.1.2}$$

where $\Phi^{-1}(\alpha)$ is the α quantile of standard normal.

It also means that if a loss random variable is normally distributed, 95% of observations will lie within 1.65 standard deviations of the mean and 99% will lie within 2.33 standard deviations.

If L is of Student t distribution $t(\nu, \mu, \sigma^2)$, then

$$VaR_\alpha = \mu + \sigma t_\nu^{-1}(\alpha), \tag{3.1.3}$$

where $t_\nu^{-1}(\alpha)$ is the α quantile of standard t with degree of freedom ν.

3.2 Data Sets

The S&P500 index is based on a portfolio of 500 different stocks in the United States. This index accounts for the 80% market capitalization of all the stocks listed on NYSE. The Dow index is based on a portfolio of 30 blue-chip stocks in the United States. The adjusted daily close prices* range from 4/18/1989 to 7/29/2005. There are 4108 observations in total. The daily close prices are converted to daily negative log returns.

*The data are pulled out from finance.yahoo.com

We have mentioned in Section 2.4 that the maximization of log likelihood assumes that the data series is approximately independent and identically distributed($i.i.d$). It is commonly believed that financial data is not $i.i.d.$ although they show little evidence of serial correlation. Squared return or absolute return series does show some evidence of serial correlation. In addition, return series is heteroscedastic, i.e., volatilities vary over time and extreme events appear in clusters.

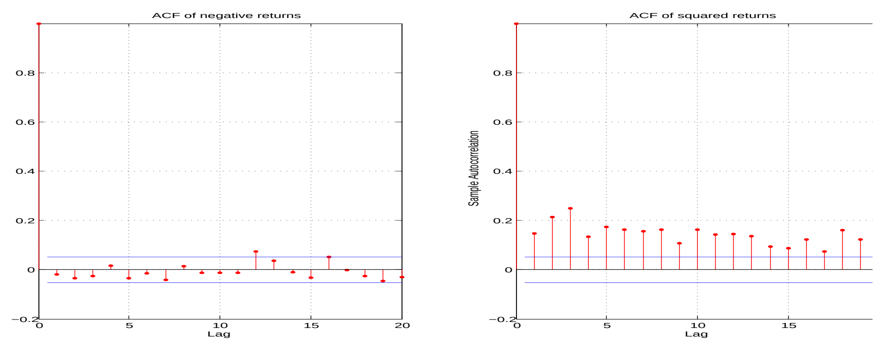

Figure 3.1: Correlograms for SP500 negative log return series

We use the most recent 1500 daily negative log return data of SP500 to plot the sample autocorrelation function(ACF). From figure 3.1, we can see that the ACF of the negative log return series shows little evidence of serial correlation, while the ACF of squared return series does show evidence of serial dependence. The $GARCH$ model can be introduced to model the persistence in the volatility.

3.3 $GARCH$ Filter

Definition 3.3.1 $ARMA$ **process with** $GARCH$ **errors**. *Let Z_t be standard white noise $SWN(0,1)$. The process (X_t) is an $ARMA(p_1, q_1)$ process with $GARCH(p_2, q_2)$ errors if it* *is covariance stationary and satisfies the following equations,*

$$X_t = \mu_t + \sigma_t Z_t, \tag{3.3.1}$$
$$\mu_t = \mu + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \theta_j (X_{t-j} - \mu_{t-j}),$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p_2} \alpha_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2$$

where $\alpha_0 > 0$,$\alpha_i \geq 0$,$\beta_j \geq 0$ *and* $\sum_{i=1}^{p_2} \alpha_i + \sum_{j=1}^{q_2} \beta_j < 1$*. The innovation* Z_t *is independent of* $(X_s)_{s<t}$*.*

McNeil et al. (2005) argued that a $GARCH(1,1)$ model with Student t innovations is enough to remove the dependence in return series. The $ARMA$ term usually is not necessary, i.e., we set $p_1 = 0$ and $q_1 = 0$. Sometimes, an filter with normal innovations is enough too. The filtered return series, which is defined to be,

$$\hat{X}_t = \frac{X_t - \mu}{\sigma_t} \tag{3.3.2}$$

should be approximately $i.i.d.$ series[†].

From figure 3.2, we can see that the ACF of both filtered return series and squared filtered return series for SP500 show little evidence of serial correlation.

3.4 Tails and Extreme Events

It is commonly believed that the return series is leptokurtic. It has more mass around the center and heavier tails than Gaussian distribution. A QQ-plot[‡] can be used to compare the empirical quantiles with quantiles of a designated distribution. Let us plot the QQ-plot against a Gaussian distribution for filtered return series of S&P500 and Dow index. From figure 3.3, we can see that both filtered returns series have heavier tails than normal distribution on both tails.

A Gaussian distribution has a kurtosis value of 3, and if the kurtosis of a distribution is greater than 3, we say this distribution is leptokurtic. The kurtosis of S&P500 return series is 3.73 and the kurtosis of Dow return series is 4.18.

Heavy tailed distributions allow more extreme events to happen. This property is valuable for risk management. Generalized hyperbolic distributions provide at least semi-heavy tails.

[†]We call garchfit in Matlab to calibrate $GARCH$ model.
[‡]We call qqplot in Matlab to plot the normal QQ-plot.

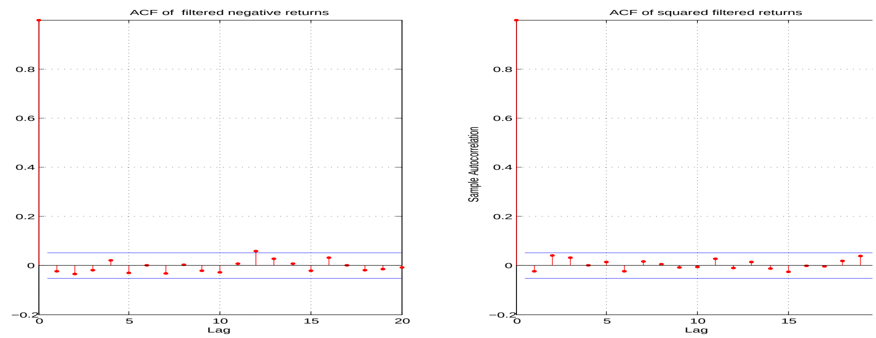

Figure 3.2: Correlograms for SP500 filtered negative log return series

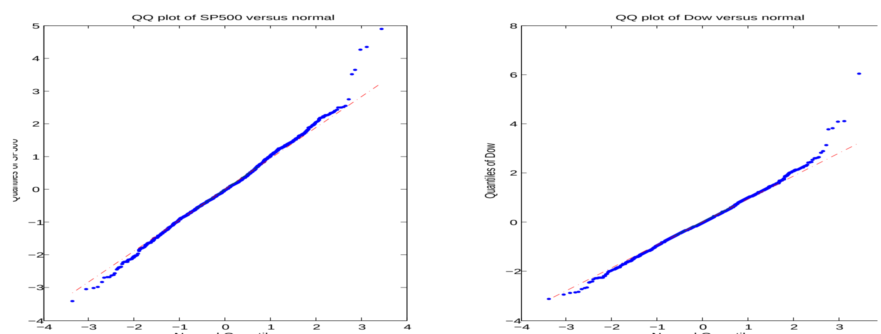

Figure 3.3: QQ-plot of S&P500 and Dow versus Normal

3.5 Density Estimation

The filtered data are now approximately *i.i.d.* so that we can calibrate the generalized hyperbolic distributions and the limiting cases by EM algorithm. We use the most recent 1500 data, which is about 6 years, to calibrate hyperbolic, NIG, skewed t, VG, Student t, and Gaussian distributions. We report the calibrated parameters and the corresponding log likelihood in table 3.1 for S&P500 index and table 3.2 for Dow index.

Table 3.1: Calibrated parameters of S&P500

Model	λ or ν	χ	ψ	μ	σ^2	γ	Likelihood
Skewed t	13.88			-0.17	0.85	0.18	-2120.69
VG	5.15		10.11	-0.20	0.97	0.23	-2120.91
NIG	-0.5	4.00	6.78	-0.18	1.29	0.29	-2120.94
Hyperbolic	1(fixed)	102.78	0.1	-0.15	0.021	0.004	-2121.83
Student t	13.95			0.03	0.86		-2121.66
Gaussian				0.04	1		-2129.53

Table 3.2: Calibrated parameters of Dow

Model	λ or ν	χ	ψ	μ	σ^2	γ	Likelihood
Skewed t	12.72			-0.19	0.83	0.19	-2117.52
VG	4.93		9.69	-0.18	0.97	0.20	-2119.06
NIG	-0.5	3.67	6.14	-0.18	1.28	0.28	-2118.48
Hyperbolic	1	101.30	0.1	-0.13	0.021	0.003	-2119.35
Student t	12.11			0.02	0.83		-2118.83
Gaussian				0.03	1		-2131.27

From the viewpoint of maximization of log likelihood, skewed t has the largest log likelihood among all the distributions tested. NIG has slightly higher log likelihood than hyperbolic. VG has second largest log likelihood in the modeling of S&P500, but third largest in the modeling of Dow. NIG and hyperbolic are well known in the modeling of financial data, however, currently skewed t are rarely known comparing with those two. In addition, skewed t has the fewest number of parameters among the four generalized hyperbolic distributions.

After we calibrate the distributions, we calculate the VaR_α[§] and list the results in table 3.3.

Table 3.3: VaR_α of sample, normal, Student t, and generalized hyperbolic distributions for S&P500 in the past 1500 days

α	sample	normal	Student t	skewed t	VG	NIG	hyperbolic	stable
0.95	1.65	1.68	1.66	1.69	1.71	1.70	1.72	1.67
0.975	2.05	2.0	2.02	2.07	2.10	2.08	2.12	2.02
0.99	2.34	2.37	2.46	2.55	2.57	2.56	2.64	2.47
0.995	2.51	2.62	2.79	2.91	2.92	2.91	2.02	2.86
0.999	4.35	3.13	3.54	3.74	3.67	3.69	3.87	4.64
0.9999	4.88	3.76	4.65	5.02	4.69	4.78	5.07	13.97
0.99999	4.88	4.31	5.86	6.90	6.90	7.05	6.82	45.13

From figure 3.4 and figure 3.6, we can see that generalized hyperbolic distributions, the empirical distribution[¶] and the stable distribution all have higher kurtosis than the normal distribution and they all have heavier tails than the normal distribution.

From figure 3.5, figure 3.7 and table 3.3, we can see more about the right tail. The empirical distribution has very poor performance on the tails. Generalized hyperbolic distributions provide a way to extrapolate outside the historical extreme events. Extrapolation is very desirable by allowing the estimation of VaR outside the historical record, and this property is invaluable for risk management applications. Generalized hyperbolic distributions have no big difference in the modeling of tails. It is commonly criticized that stable distribution has too fat tails. Skewed t has heavier tails than Student t because of the skewness parameters.

From figure 3.8, figure 3.9, figure 3.10, and figure 3.11, we can see that QQ-plots also show that generalized hyperbolic distributions have heavier tails than the normal distribution.

3.6 Backtesting of VaR

After we calibrate the filtered negative log return series using generalized hyperbolic distributions, we can calculate the α quantile, $z_\alpha = F^{-1}(\alpha)$ for filtered negative return series, where F is some distribution function. Suppose that we are standing at time t, we

[§]For generalized hyperbolic distributions, we get VaR_α by finding y from $\int_{-\infty}^{y} f(x)dx - \alpha = 0$.

[¶]We call ksdensity in Matlab to plot the empirical density.

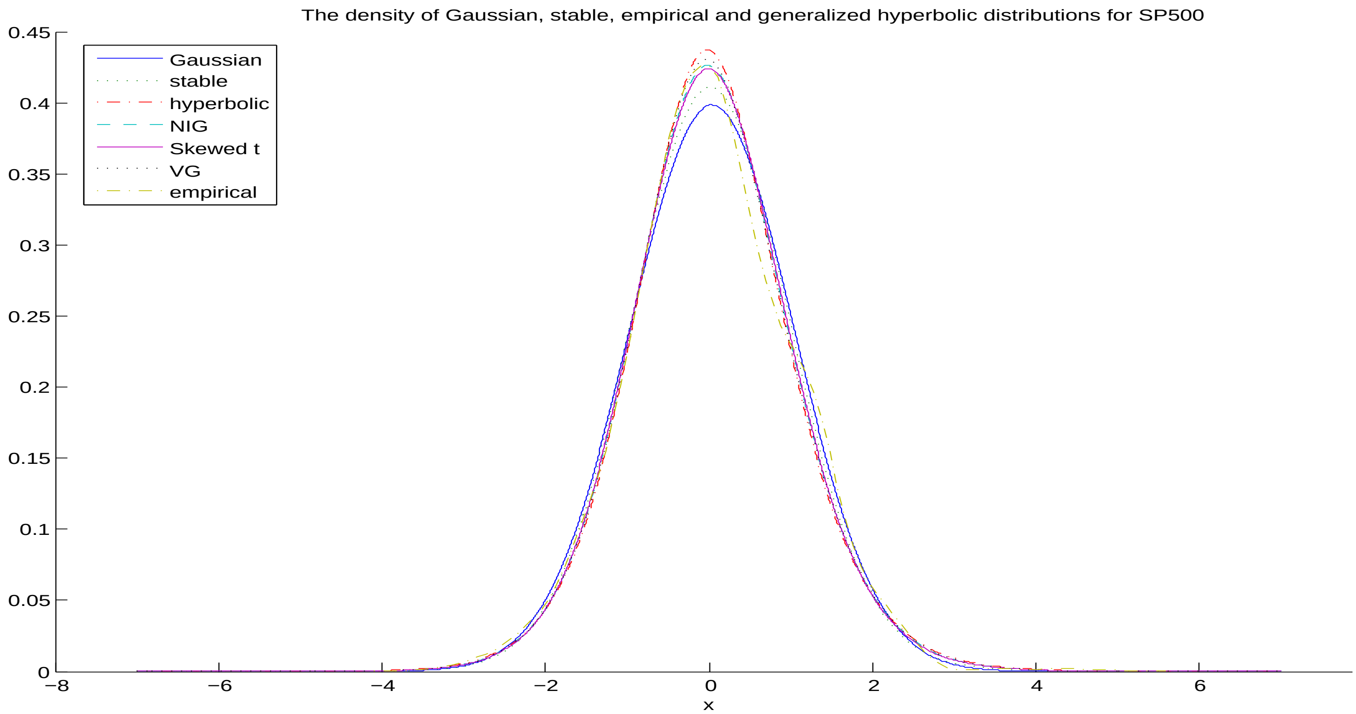

Figure 3.4: The density of Gaussian, stable, empirical and generalized hyperbolic distributions for S&P500

then de-filter the value at risk(VaR) of filtered return series to get the estimation of VaR of negative log return series at time $t+1$.

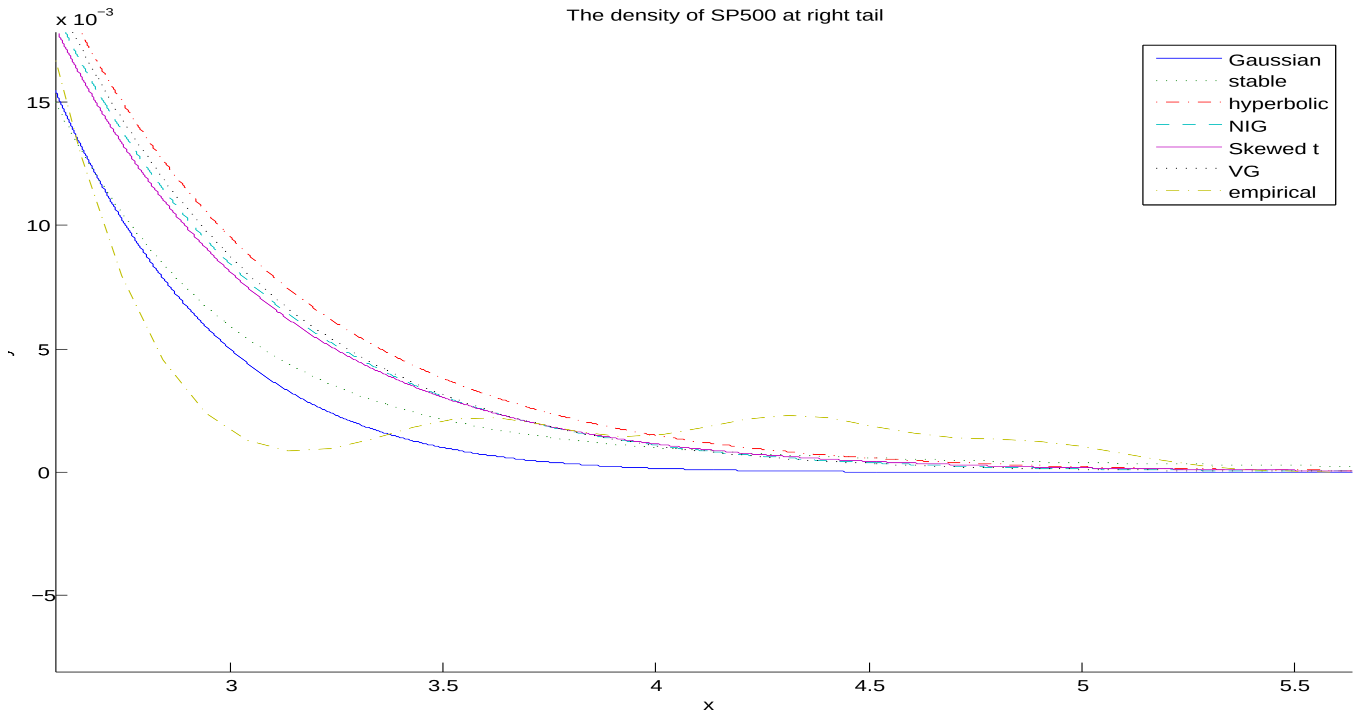

Figure 3.5: The density of Gaussian, stable, empirical and generalized hyperbolic distributions for S&P500 at right tail

$$V\hat{aR}_{\alpha}(X_{t+1}|\mathcal{F}_t) = \hat{\sigma}_{t+1}z_{\alpha} + \mu, \tag{3.6.1}$$

where σ_{t+1} can be forecasted by using equation 3.3.1 and it is known at time t.

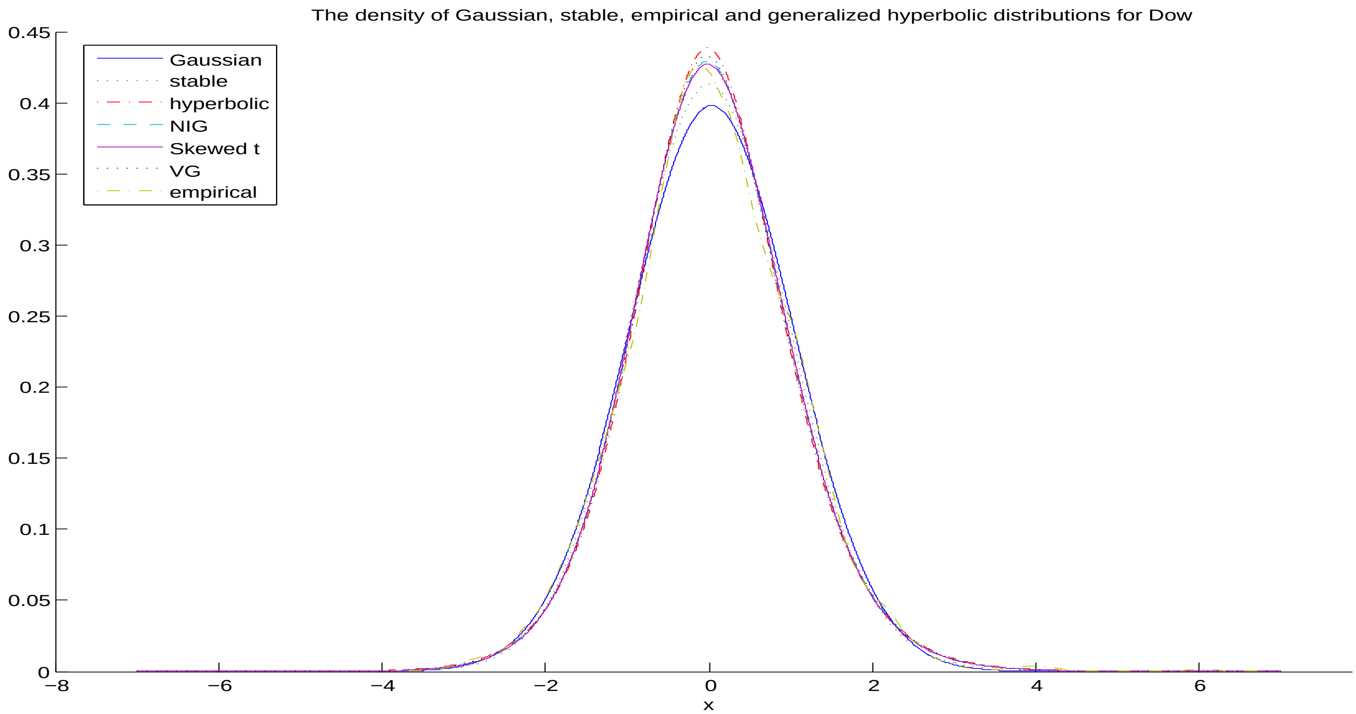

Figure 3.6: The density of Gaussian, stable, empirical and generalized hyperbolic distributions for Dow

We have 4108 observations for both S&P and Dow index. We use the most recent $N = 3001$ observations to backtest the VaR violations. For each day, we use previous 1000 observations to train the generalized hyperbolic distributions. We recalibrate the

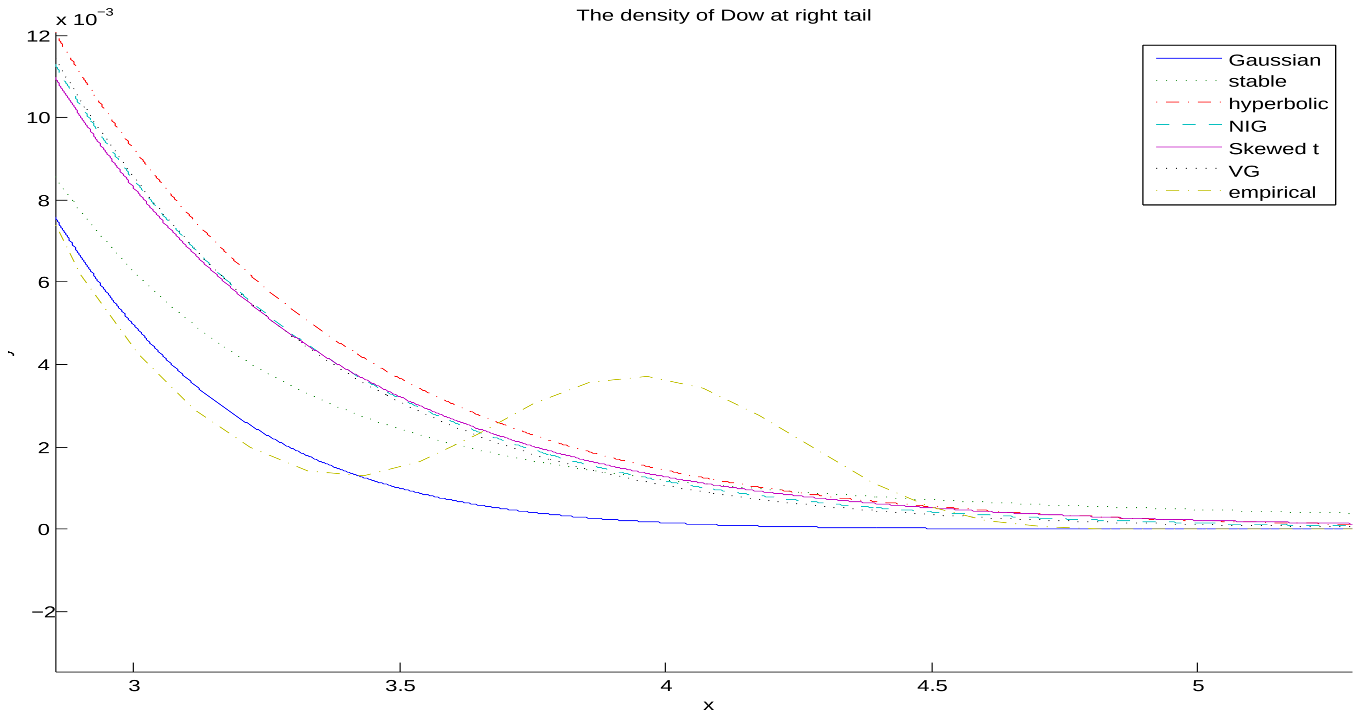

Figure 3.7: The density of Gaussian, stable, empirical and generalized hyperbolic distributions for Dow at right tail

model at each day by initializing the parameters using last day's and recalibrate the model by initializing the parameters using new set of parameters every 400 days. In this way, the calibration is always online, and the recalibration every 400 days by initializing the

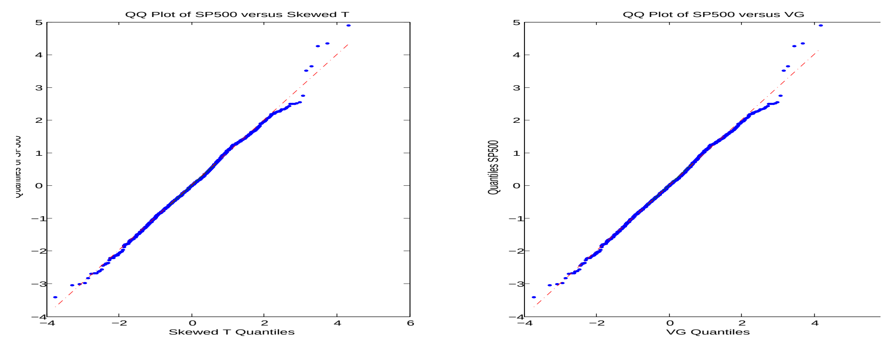

Figure 3.8: QQ-plot of S&P500 versus Skewed t and VG

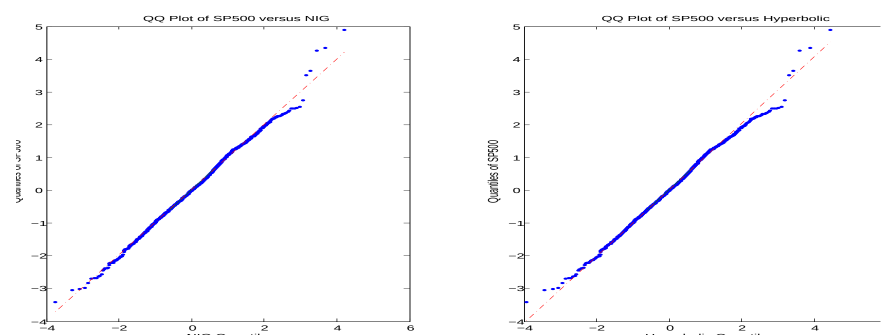

Figure 3.9: QQ-plot of S&P500 versus NIG and hyperbolic

parameters using brand new parameters avoids over estimation. At time t+1, where t is from 1000 from 3999, we use $(x_{t-1000+1}, \ldots, x_{t-1}, x_t)$ to calibrate the generalized hyperbolic

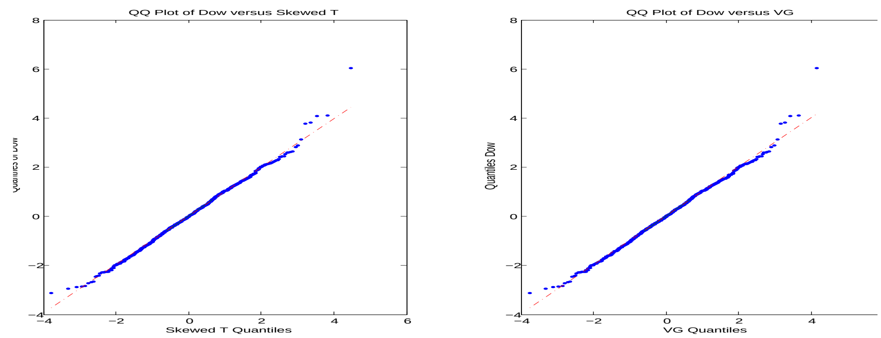

Figure 3.10: QQ-plot of Dow versus Skewed t and VG

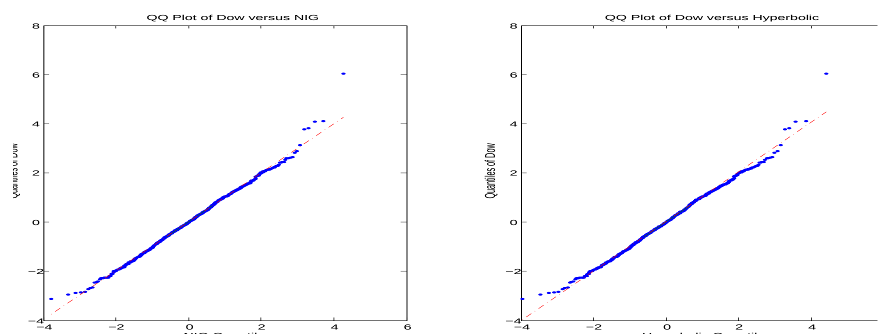

Figure 3.11: QQ-plot of Dow versus NIG and hyperbolic

distributions and estimate $VAR_\alpha(X_{t+1}|\mathcal{F}_t)$ for $\alpha = 0.95$,$\alpha = 0.975$,$\alpha = 0.99$, and $\alpha = 0.995$. A violation occurs if $x_{t+1} > V\hat{A}R_\alpha(X_{t+1}|\mathcal{F}_t)$. The number of total violations during those

N tests is denoted to be n. The actual violation probability is n/N, while the expected violation probability, q, should be 0.05, 0.025, 0.01 and 0.005 respectively.

To evaluate the VaR backtesting, we use the likelihood ratio statistic by Kupiec(1995). The null hypothesis is that the expected violation probability is equal to q. Under null hypothesis, the likelihood ratio given by

$$-2[(N-n)\log(1-q)+n\log(q)]+2[(N-n)\log(1-n/N)+n\log(n/N)], \quad (3.6.2)$$

is asymptotically $\chi^2(1)$ distributed.

We list the results of VaR backtesting in table 3.4 for S&P500 index and in table 3.5 for Dow index. We calculate the actual violation probability at level q, where the expected violation probability, q, is 0.05, 0.025, 0.01 and 0.005 respectively, and its corresponding p-value[‖] for the likelihood ratio test. We backtest normal distribution and four generalized hyperbolic distributions in this paper. we can see that all generalized hyperbolic distributions pass the test. The normal distribution fails the test at 0.005 level and 0.01 level for both S&P500 and Dow index and at 0.025 level for S&P500[**].

Table 3.4: VaR violation backtesting for S&P500

Model	0.05	p	0.025	p	0.01	p	0.005	p
Normal	0.0477	0.5547	0.0313	**0.0324**	0.0177	**0.0001**	0.0117	**0.0000**
Skewed t	0.0487	0.7365	0.0260	0.7274	0.0093	0.7105	0.0043	0.5962
VG	0.0460	0.3085	0.0257	0.8159	0.0103	0.8552	0.0053	0.7979
NIG	0.0467	0.3971	0.0253	0.9071	0.0093	0.7105	0.0047	0.7935
Hyperbolic	0.0453	0.2337	0.0240	0.7240	0.0097	0.8536	0.0047	0.7935

3.7 Conclusion

The normal distribution tends to underestimate the risk since it is a thin tailed distribution. Generalized hyperbolic distributions have semi-heavy tails so that they may be good candidates for risk management. We use a $GARCH$ model to filter the negative return series to get $i.i.d.$ filtered negative return series and forecast the volatility. After we get $i.i.d.$

[‖] We call CHIDIST(x,1) in Excel to calculate the p-value, where x is the value of likelihood ratio statistic.
[**] When p value is less than 0.05, we reject the null hypothesis.

Table 3.5: VaR violation backtesting for Dow

Model	0.05	p	0.025	p	0.01	p	0.005	p
Normal	0.0487	0.7365	0.0303	0.0701	0.0160	**0.0024**	0.0113	**0.0000**
Skewed t	0.0480	0.6130	0.0247	0.9067	0.0100	1.0000	0.0063	0.3202
VG	0.0463	0.3511	0.0240	0.7240	0.0107	0.7166	0.0073	0.0904
NIG	0.0470	0.4465	0.0240	0.7240	0.0100	1.0000	0.0070	0.1431
Hyperbolic	0.0460	0.3085	0.0240	0.7240	0.0093	0.7105	0.0067	0.2183

filtered negative return series, we can calibrate the generalized hyperbolic distributions and calculate the α quantile. Using the forecasted volatility and α quantile for filtered negative return series, we can restore the VaR_α for negative return series. We backtest VaR based on generalized hyperbolic distributions and normal distribution and find that all generalized hyperbolic distributions pass the VaR test while normal distribution fails at 99% VaR for both Dow and S&P500 index, even at 97.5% level for S&P500.

The use of the skewed t distribution is rarely known. Actually, it has the fewest number of parameters among all generalized hyperbolic distributions and a fast calibration algorithm. In addition, it has the largest log likelihood among all generalized hyperbolic distributions, Student t, and Gaussian distributions. We will focus on this distribution later.

CHAPTER 4

A $GARCH$-Skewed t-ES Portfolio Optimization Model

Portfolio optimization is based on trading off risk and return. For this purpose one needs to employ some precise concept of "risk". Markowitz (1952) suggested using the standard deviation of portfolio return as a risk measure, and, thinking of returns as normally distributed, described the efficient frontier, which is composed of fully invested portfolios having minimum risk for a given specified return. This concept has been extremely valuable in portfolio management because any rational portfolio manager will always choose to invest on this frontier.

However, using standard deviation as the risk measure has the drawback that it is generally insensitive to extreme events, and sometimes these are of most interest to the investor.

Value at Risk (VaR) can describe more about extreme events, but it can not aggregate risk in the sense of being subadditive on portfolios (which means risk is diversified). This is a well-known difficulty that is addressed by the concept of a "coherent risk measure" in the sense of Artzner, Delbaen, Eber, and Heath(1999). A popular example of a coherent risk measure is the expected shortfall (ES), though VaR is still more commonly seen in practice.

The construction of an efficient frontier – portfolios with minimum risk for a given return – depends on two inputs: the choice of risk measure (such as standard deviation, VaR, or ES), and the probability distribution used to model returns.

It turns out, by a result of Embrechts, McNeil, and Straumann (2001), that when the underlying distribution is Gaussian – or more generally any "elliptical" distribution – no matter what positive homogeneous and translation invariant risk measure(such as VaR, or ES), no matter what confidence level, the optimized portfolio composition given a certain return will be the same as the traditional Markowitz style portfolio composition obtained by minimizing standard deviation. Only a difference in distribution leads to different portfolio compositions.

As mentioned in last chapter, portfolio managers can not neglect the deviation of financial returns series from a multivariate normal distribution. Other heavy tailed elliptical distributions, such as Student t and symmetric generalized hyperbolic distributions, and non-elliptical distributions, such as the skewed t distribution, can be used to model financial returns series. If the underlying is Gaussian distributed, then the portfolio return is Gaussian distributed too. More generally, if the underlying is generalized hyperbolic distributed, then the portfolio return remains in the same sub-family of generalized hyperbolic distributions since they are closed under linear transformations. The usual parametrization of generalized hyperbolic distributions generally does not exhibit this property. In addition, for Gaussian, Student t and symmetric generalized hyperbolic distributions, the portfolio variance is in quadratic form so that it is easy to minimize.

For non-elliptical distributions, different risk measures, or the same risk measure with different confidence levels, lead to differing portfolio compositions given a fixed return. Rockafellar and Uryasev (1999) showed that the minimization of ES does not require knowing VaR first, and construct a new objective function. By minimizing this new objective function, we can get VaR and ES. We carry this out and use Monte Carlo simulation to approximate that new objective function by sampling the multivariate distributions. This allows us to construct efficient frontiers for a variety of distributions.

From the last chapter, we can see that the skewed t distribution has the largest log likelihood and the fewest number of parameters among the four generalized hyperbolic distributions tested when we model the univariate Dow and S&P 500 index daily close price series. In addition, skewed t has a fast calibration algorithm among all the tested generalized hyperbolic distributions. It is natural to extend the univariate application of skewed t to the multivariate case. To our knowledge, this is first use of skewed t distribution in portfolio optimization.

This chapter is organized as following. We consider an equity portfolio of 5 stocks chosen from components of the Dow index and filter the data to get *i.i.d.* filtered returns series in Section 4.1. After we get the *i.i.d.* filtered returns series, we fit the five dimensional data by normal inverse Gaussian(NIG), hyperbolic, skewed t and variance gamma (VG) in Section 4.2 using EM algorithm and find that skewed t has the largest log likelihood and the fewest number of parameters among the four tested generalized hyperbolic distributions . In Section 4.3, we discuss coherent risk measure: ES, and portfolio optimization theory under elliptical distributions. In Section 4.4, we plot the normal frontier, Student t frontier, and skewed t frontier. One consequence we describe is that the usual Gaussian efficient frontier is actually unreachable if we believe that returns are Student t or skewed t distributed.

4.1 Data Sets

We construct the portfolio by choosing 5 stocks from the components of the Dow index. They are WALT DISNEY, EXXON MOBIL, PFIZER, ALTRIA GROUP and INTEL*. We use the adjusted close data ranged from 7/1/2002 to 08/04/2005. The daily close data are converted to log return. Figure 4.1 illustrates the relative price movements of each index using most recent 750 returns. The initial price of each stock is rescaled to one to facilitate the comparison of relative performance.

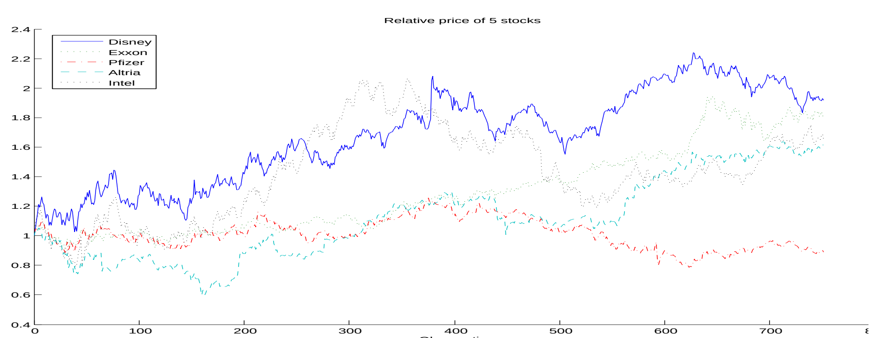

Figure 4.1: Relative price of 5 stocks

From figure 4.2, we can see that squared returns series show some evidence of serial correlation as before.

*The data set is obtained from finance.yahoo.com.

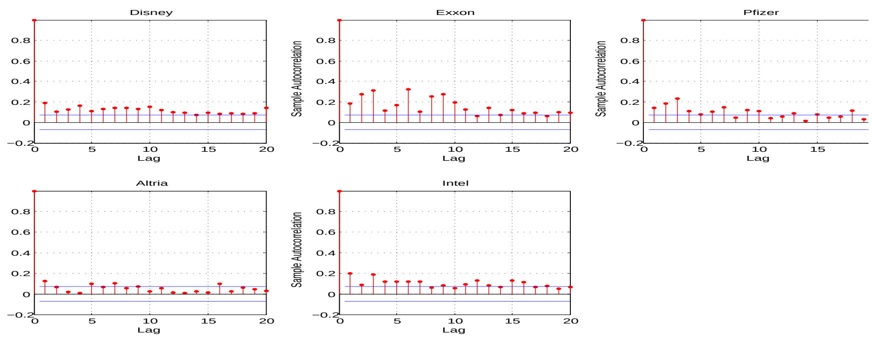

Figure 4.2: Correlograms of squared log return series for 5 stocks

4.1.1 $GARCH$ Filter

A $GARCH(1,1)$ model with Gaussian innovation is used to remove the dependence. We set the constant μ in equation 3.3.1 to be zero. From figure 4.3, we can see that squared filtered returns series show no evidence of serial correlation. From figure 4.4, we can see that heteroscedasticity clearly exists in 5 stocks.

4.2 Multivariate Density Estimation

After we get the approximately $i.i.d.$ training data, we can estimate the multivariate density. From QQ-plots versus normal for those 5 stocks in figure 4.5, we can see that normal distribution has very thin tails so that it may not be used to handle the risk management. We use generalized hyperbolic distributions to model the multivariate density. Table

Model	Normal	Student t	Skewed t	VG	Hyperbolic	NIG
Log Likelihood	-5095.01	-4877.76	-4873.91	-4901.74	-4891.46	-4884.15

Table 4.1: Log likelihood of estimated multivariate density

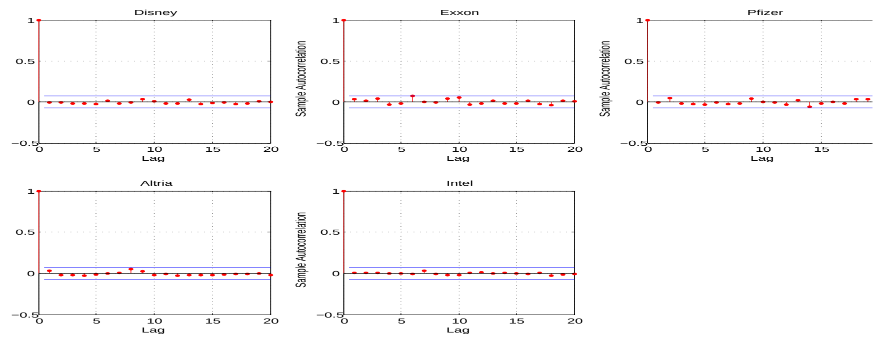

Figure 4.3: Correlograms of squared filtered log return series for 5 stocks

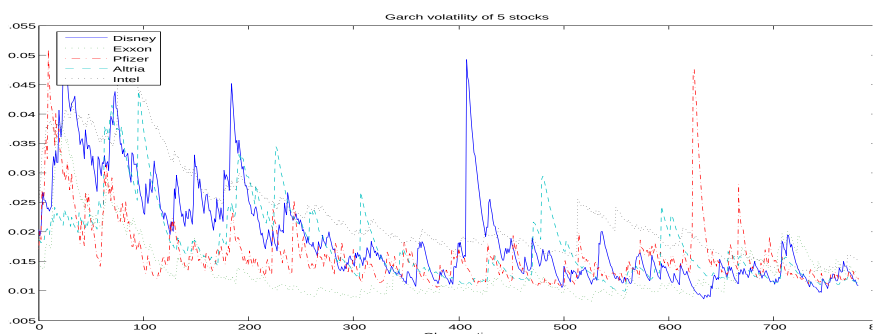

Figure 4.4: $GARCH$ volatility of log return series for 5 stocks over time

4.1 shows the maximized log likelihood. It shows again that all generalized hyperbolic distributions have higher log likelihood than normal distribution and skewed t has the highest

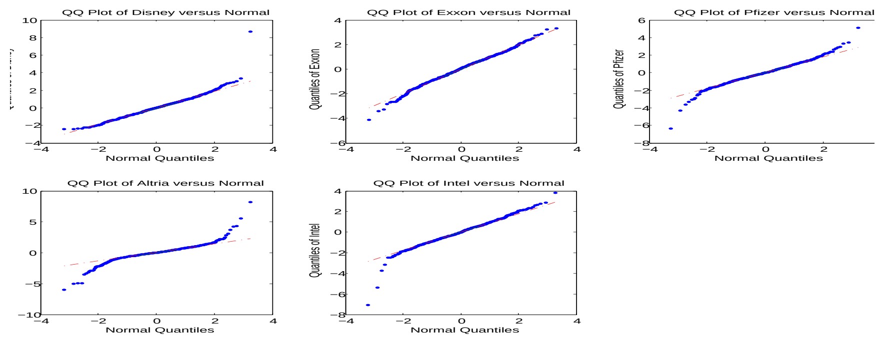

Figure 4.5: QQ-plot versus normal distribution of 5 stocks

log likelihood. In the sequel, we will largely use skewed t in the multivariate modeling and correlation modeling.

4.3 Risk Measure and Portfolio Optimization

Suppose $\boldsymbol{\omega}^T = (\omega_1, \cdots, \omega_d)$ is the capital amount invested on each risky security in a portfolio, and $\mathbf{X}^T = (X_1, \cdots, X_d)$ is the return of each risky security. Let $L(\boldsymbol{\omega}, \mathbf{X}) = -\sum_{i=1}^{d} \omega_i X_i = -\boldsymbol{\omega}^T \mathbf{X}$ denote the loss of of this portfolio over a fixed time interval Δ and F_L is its distribution function. Usually, the time interval Δ is one or ten days for equity portfolio management.

If $\mathbf{X}$ is of normal distribution denoted by $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, then

$$L \sim N_1(\text{-}\boldsymbol{\omega}^T \boldsymbol{\mu}, \boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}) \tag{4.3.1}$$

If $\mathbf{X}$ is of Student t distribution denoted by $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, then

$$L \sim t_1(\nu, -\boldsymbol{\omega}^T \boldsymbol{\mu}, \boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}) \tag{4.3.2}$$

Both normal distribution and Student t distribution are elliptical distributions.

If $\mathbf{X}$ is of skewed t distribution denoted by $\mathbf{X} \sim SkewedT_d(\nu, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$, then

$$L \sim SkewedT_1(\nu, -\boldsymbol{\omega}^T\boldsymbol{\mu}, \boldsymbol{\omega}^T\Sigma\boldsymbol{\omega}, -\boldsymbol{\omega}^T\boldsymbol{\gamma}) \tag{4.3.3}$$

Skewed t distribution is not elliptical when $\boldsymbol{\gamma} \neq \mathbf{0}$.

We have talked about two risk measures before, standard deviation and value at risk. Standard deviation is still commonly used in Markowitz style portfolio management. VaR is the standard risk measure used by regulators and investment banks. Standard deviation risk measure is criticized by its inability to describe the rare events and VaR is criticized by its inability to aggregate risk. Standard deviation is not a coherent risk measure. VaR is a coherent measure if the underlying distribution is elliptical.

Definition 4.3.1 Expected Shortfall(ES). *For a continuous loss distribution with $\int_{\mathbb{R}} |l| dF_L(l) < \infty$, the ES_α at confidence level $\alpha \in (0,1)$ for loss L of a security or a portfolio is defined to be*

$$ES_\alpha = E(L|L \geq VaR_\alpha) = \frac{\int_{VaR_\alpha}^{\infty} l dF_L(l)}{1-\alpha}. \tag{4.3.4}$$

If L is of normal distribution $N(\mu, \sigma^2)$, then

$$ES_\alpha = \mu + \sigma\frac{\psi(\Phi^{-1}(\alpha))}{1-\alpha}, \tag{4.3.5}$$

where ψ is the density of standard normal distribution and $\Phi^{-1}(\alpha)$ is the α quantile of standard normal.

If L is of Student t distribution $t(\nu, \mu, \sigma^2)$, then

$$ES_\alpha = \mu + \sigma\frac{f_\nu(t_\nu^{-1}(\alpha))}{1-\alpha}\left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu - 1}\right), \tag{4.3.6}$$

where f_ν is density function of standard t with degree of freedom ν and $t_\nu^{-1}(\alpha)$ is the α quantile of standard t with degree of freedom ν.

For skewed t, there is no closed solution for or VaR or ES. To calculate VaR or ES, we can use numerical integration and a zero-finder routine. We can also use Monte Carlo simulation by using the following definition to get ES:

Definition 4.3.2 Expected Shortfall(ES). *If the density function of $\mathbf{X}$ is $f(\mathbf{x})$, the expected shortfall at confidence level $\alpha \in (0,1)$ for loss L of a security or a portfolio is defined to be*

$$ES_\alpha = E(L|L \geq VaR_\alpha) = \frac{\int I_{\{-(\boldsymbol{\omega}^T\mathbf{x}) \geq VaR_\alpha\}} - (\boldsymbol{\omega}^T\mathbf{x}) f(\mathbf{x}) d\mathbf{x}}{1-\alpha}. \tag{4.3.7}$$

Expected shortfall is a coherent risk measure.

Definition 4.3.3 Coherent Risk Measure(Artzner, Delbaen, Eber and Heath(1999)). *A real valued function ρ of a random variable is a coherent risk measure if it satisfies the following properties,*

1. **Subadditivity**. *For any two random variables X and Y, $\rho(X+Y) \leq \rho(X) + \rho(Y)$.*
2. **Monotonicity**. *For any two random variables $X \geq Y$, $\rho(X) \geq \rho(Y)$.*
3. **Positive homogeneity**. *For $\lambda \geq 0$, $\rho(\lambda X) = \lambda \rho(X)$.*
4. **Translation invariance**. *For any $a \in \mathbb{R}$, $\rho(a + X) = a + \rho(X)$.*

Proposition 4.3.4 Portfolio Optimization under Elliptical Distributions. (Embrechts, McNeil and Straumann(2001)). *Suppose $\mathbf{X}$ is of elliptical distribution with finite variance for all univariate marginals. Let*

$$\mathcal{P} = \{Z = \sum_{i=1}^{d} \omega_i X_i | \omega_i \in \mathbb{R}\}$$

be the set of all linear portfolios. Then:

1. **Subadditivity of VaR.** *For any two portfolios Z_1 and $Z_2 \in \mathcal{P}$, and $0.5 \leq \alpha < 1$,*

$$VaR_\alpha(Z_1 + Z_2) \leq VaR_\alpha(Z_1) + VaR_\alpha(Z_2)$$

2. **Equivalence of variance and any other positive homogeneous risk measure**. *Let ρ be a risk measure depending only on the distribution of a random variable X that is positively homogeneous. Then for Z_1 and $Z_2 \in \mathcal{P}$,*

$$\rho(Z_1 - E(Z_1)) \leq \rho(Z_2 - E(Z_2)) \Longleftrightarrow \sigma^2_{Z_1} \leq \sigma^2_{Z_2}$$

3. **Markowitz risk minimizing portfolio**. *Let ρ be as in (2) and also translation invariant, and let*

$$\mathcal{Q} = \{Z = \sum_{i=1}^{d} \omega_i X_i | \omega_i \in \mathbb{R}, \sum_{i=1}^{d} \omega_i = 1, E(Z) = r\}$$

be the subset of $\mathcal{P}$ with given expected return r , then

$$argmin_{Z \in \mathcal{Q}} \rho(Z) = argmin_{Z \in \mathcal{Q}} \sigma^2_Z.$$

This proposition shows that if we assume that the underlying distribution is elliptical, then the Markowitz mean variance optimized portfolio, for a given return, will be the same as the optimized portfolio by minimizing any other translation invariant and positively homogeneous risk measure, such as VaR or ES. Only a difference in distribution will lead to a different portfolio allocation at a given return. The thin tailed normal distribution is not required in portfolio management. Portfolio managers may consider heavy tailed elliptical distributions such as Student t, and symmetric generalized hyperbolic distributions etc.

For some elliptical distributions such as the normal distribution, Student t distribution, and symmetric generalized hyperbolic distributions, we can get the frontier by minimizing the portfolio variance. The portfolio variance is in quadratic form so that it is easy to optimize. We can also optimize portfolio VaR and ES directly if they are in closed form. After we get the optimized portfolio, we can get other portfolio risk measures. If we try to minimize VaR or ES instead of variance, for a give return, the portfolio composition is the same.

The generalized hyperbolic distribution is not elliptical if $\boldsymbol{\gamma} \neq \mathbf{0}$. From corollary 2.2.4 and equation 2.2.8, we can see that the portfolio variance is not in quadratic form anymore. We will turn to Monte Carlo simulation to minimize a coherent risk measure. Specifically, we minimize ES at confidence level α by sampling the multivariate distribution of returns. Without doubt, we can also try this method for all elliptical distributions mentioned above.

From definition 4.3.2, we can rewrite the definition of expected shortfall as following,

$$ES_\alpha = VaR_\alpha + \frac{\int [-(\boldsymbol{\omega}^T \mathbf{x}) - VaR_\alpha]^+ f(\mathbf{x}) d\mathbf{x}}{1-\alpha},$$

where $[x]^+ := max(x, 0)$.

We get a new objective function by replacing VaR by p,

$$F_\alpha(\boldsymbol{\omega}, p) = p + \frac{\int [-(\boldsymbol{\omega}^T \mathbf{x}) - p]^+ f(\mathbf{x}) d\mathbf{x}}{1-\alpha}. \tag{4.3.8}$$

Rockafellar and Uryasev(2001) showed that

$$min_{p\in\mathbb{R}} F_\alpha(\boldsymbol{\omega}, p) = min_{(\boldsymbol{\omega},p)\in\mathbb{R}^d\times\mathbb{R}} F_\alpha(\boldsymbol{\omega}, p).$$

If the p minimizing equation 4.3.8 with respect to p is unique, then by minimizing equation 4.3.8 with respect to $(\boldsymbol{\omega}, p)$, we obtain $(\boldsymbol{\omega}^*, p^*)$, where $\boldsymbol{\omega}^*$ is the optimized portfolio composition and $p*$ is the corresponding portfolio's VaR at confidence level α.

We can sample the multivariate density by Monte Carlo simulation to estimate the $F_\alpha(\boldsymbol{\omega}, p)$ by the following

$$\hat{F_\alpha}(\boldsymbol{\omega}, p) = p + \frac{\sum_{k=1}^{n}[-(\boldsymbol{\omega}^T \mathbf{x}_k) - p]^+}{n(1-\alpha)}, \tag{4.3.9}$$

where $\mathbf{x}_k$ is the k-th sample from some distribution and n is the number of samples.

4.4 Efficient Frontier Analysis

Suppose that we are standing at Aug 4,2005, i.e. the last date in our data set and the holding period is one day. 750 sample data are used in the calibration. The one day ahead forecasted $GARCH$ volatilities for all the stocks are $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_d)^T$ at that date. The weight constraint condition is written as

$$\sum_{i=1}^{d} \omega_i = 1, \tag{4.4.1}$$

where we assume the initial capital is 1 and ω_i is the capital invested in risky stock i. If no short sales allowed, we set $\omega_i > 0$. We suppose short sales are allowed.

Suppose that the calibrated expected return of stock i is $\hat{\mu}_i$, then the de-filtered forecasted expected return is $\mu_i = \sigma_i \hat{\mu}_i$ and let

$$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_d)^T,$$

then the expected portfolio return † is $\boldsymbol{\omega}^T \boldsymbol{\mu}$. We set the expected portfolio return to be a constant series,

$$\boldsymbol{\omega}^T \boldsymbol{\mu} = c. \tag{4.4.2}$$

The objective function to be optimized can be portfolio variance, VaR, ES and equation 4.3.9 etc. The constraints are equation 4.4.1 and equation 4.4.2.

In the following, we construct normal, Student t, and skewed t efficient frontiers versus different risk measures for the 5 stocks chosen from Dow index components.

Normal Frontier

†If the holding period is 1 day, there is no too much difference between logarithmic return and arithmetic return. If holding period is over a certain days, to get the portfolio return, it is suggested to convert the logarithmic returns into arithmetic returns and get the portfolio arithmetic return by summing the weighted arithmetic returns, and finally convert the portfolio arithmetic return back to logarithmic return.

Table 4.2 shows the expected log return for the filtered data and the $GARCH$ volatility on Aug 4, 2005. Table 4.3 shows the covariance matrix(upper triangular), the variance(diagonal) and correlation matrix(lower triangular) for the 5 stocks.[‡]

Table 4.2: Expected filtered log return and one day ahead forecasted $GARCH$ volatility on 08/04/2005

Stock	Disney	Exxon	Pfizer	Altria	Intel
Expected filtered return	0.040	0.073	-0.015	0.039	0.027
$GARCH$ volatility	0.0107	0.0128	0.0130	0.0113	0.0156

Table 4.3: Covariance and correlation matrix of normal distribution for filtered returns series: the diagonal are variance, the upper triangular is covariance and the lower triangular is correlation

Stock	Disney	Exxon	Pfizer	Altria	Intel
Disney	1.009	0.372	0.340	0.191	0.428
Exxon	0.367	1.015	0.364	0.199	0.309
Pfizer	0.337	0.359	1.008	0.217	0.302
Altria	0.189	0.197	0.215	1.009	0.171
Exxon	0.420	0.303	0.297	0.168	1.027

We try to minimize the portfolio variance [§]and 99% ES[¶] to get the normal frontier. We also use Monte Carlo simulation to minimize 99% ES[‖]. We generate 1,000,000 samples in the Monte Carlo simulation. Table 4.4 shows the portfolio compositions and the corresponding standard deviation, 99% VaR and 99% ES. These three methods all get the same portfolio composition for a given return. Figure 4.6 shows the three efficient frontiers based on three different methods using 99% ES as the risk measure are the same. In addition, no matter

[‡]The expected return $\hat{\boldsymbol{\mu}}$ and covariance matrix $\hat{\Sigma}$ are calibrated using filtered return. We need to restore the original expected return $\boldsymbol{\mu}$ and covariance matrix Σ by $\mu_i = \hat{\mu_i}\sigma_i$ and $\Sigma = A\hat{\Sigma}A$, where A=Diag($\boldsymbol{\sigma}$). Furthermore, for the negative return series, the mean of the loss is $-\boldsymbol{\mu}$ and the covariance matrix is Σ.

[§]We call quadprog in Matlab.

[¶]We call fmincon in Matlab, where the function to be optimized is the explicit form of 99% ES.

[‖]We call fmincon in Matlab, where the function to be optimized is the Monte Carlo simulation of 99% ES.

what confidence level α we choose, the portfolio composition at a given return will not change.

Table 4.4: Portfolio composition and corresponding standard deviation, 99%VaR and 99%ES for normal frontier

Return	deviation	99%VaR	99% ES	Disney	Exxon	Pfizer	Altria	Intel
0	0.0096	0.0223	0.0256	0.319	-0.206	0.528	0.320	0.040
0.0002	0.0084	0.0194	0.0222	0.318	-0.038	0.344	0.333	0.043
0.0004	0.0079	0.0180	0.0207	0.318	0.131	0.161	0.345	0.045
0.0006	0.0082	0.0186	0.0214	0.317	0.300	-0.023	0.358	0.048
0.0008	0.0093	0.0209	0.0241	0.317	0.468	-0.206	0.371	0.050
0.001	0.0109	0.0244	0.0281	0.316	0.637	-0.390	0.384	0.052
0.0012	0.0129	0.0287	0.0331	0.316	0.806	-0.573	0.397	0.055
0.0014	0.0150	0.0335	0.0386	0.316	0.974	-0.757	0.409	0.057
0.0016	0.0173	0.0386	0.0444	0.315	1.143	-0.940	0.422	0.060
0.0018	0.0196	0.0438	0.0505	0.315	1.312	-1.124	0.435	0.062
0.002	0.0220	0.0492	0.0567	0.314	1.480	-1.307	0.448	0.065

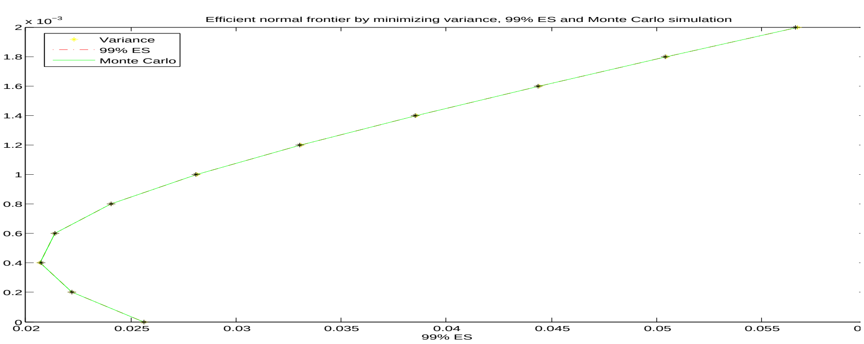

Figure 4.6: Markowitz efficient frontier

Student t Frontier

Table 4.5 shows the expected log return for the filtered data using Student t distribution. The degree of freedom is 5.87. Table 4.6 shows the dispersion matrix(upper triangular), and correlation matrix(lower triangular) for the 5 stocks**.

Table 4.5: Expected filtered log return

Stock	Disney	Exxon	Pfizer	Altria	Intel
Expected filtered return	0.015	0.077	-0.018	0.069	0.030
$GARCH$ volatility	0.0107	0.0128	0.0130	0.0113	0.0156

Table 4.6: Dispersion and correlation matrix of Student t distribution: the diagonal are diagonal of dispersion matrix, the upper triangular is dispersion matrix and the lower triangular is correlation matrix

Stock	Disney	Exxon	Pfizer	Altria	Intel
Disney	0.709	0.268	0.267	0.159	0.332
Exxon	0.363	0.771	0.274	0.170	0.244
Pfizer	0.378	0.373	0.702	0.155	0.250
Altria	0.265	0.271	0.259	0.511	0.138
Exxon	0.460	0.324	0.349	0.225	0.734

We also have three methods to get the Student t frontier. The methods are the same as the methods used in obtaining normal frontier. Table 4.7 shows the portfolio compositions and the corresponding standard deviation, 99% VaR and 99% ES. These three methods all get the same results. In addition, no matter what confidence level α we choose, the portfolio composition at a given return will not change. From this table, we can see that the portfolio compositions are different from normal frontier. Figure 4.8 shows these three methods are equivalent.

From table 4.4, table 4.7 and figure 4.8, we can see that normal frontier can not be reached if the risk measure is 99% ES and if we believe the true distribution is Student t. Multivariate Student t distribution fits the returns data better than multivariate normal

**The expected return $\hat{\boldsymbol{\mu}}$ and dispersion matrix $\hat{\Sigma}$ are calibrated using filtered return. We need to restore the original expected return $\boldsymbol{\mu}$ and dispersion matrix Σ by $\mu_i = \hat{\mu_i}\sigma_i$ and $\Sigma = A\hat{\Sigma}A$, where A=Diag($\boldsymbol{\sigma}$). Furthermore, for the negative return series, the mean of the loss is $-\boldsymbol{\mu}$ and the dispersion matrix is Σ.

Table 4.7: Portfolio composition and corresponding standard deviation, 99%VaR and 99%ES for Student t frontier

Return	deviation	99%VaR	99% ES	Disney	Exxon	Pfizer	Altria	Intel
0	0.0095	0.0245	0.0316	0.494	-0.153	0.447	0.247	-0.035
0.0002	0.0086	0.0218	0.0281	0.410	-0.048	0.315	0.336	-0.014
0.0004	0.0080	0.0203	0.0262	0.326	0.057	0.184	0.425	0.008
0.0006	0.0081	0.0201	0.0261	0.242	0.162	0.052	0.515	0.030
0.0008	0.0086	0.0214	0.0278	0.158	0.267	-0.080	0.604	0.051
0.001	0.0097	0.0238	0.0310	0.074	0.371	-0.211	0.693	0.073
0.0012	0.0110	0.0271	0.0352	-0.010	0.476	-0.343	0.782	0.094
0.0014	0.0126	0.0309	0.0402	-0.094	0.581	-0.474	0.871	0.116
0.0016	0.0143	0.0352	0.0457	-0.178	0.686	-0.606	0.961	0.138
0.0018	0.0161	0.0396	0.0515	-0.262	0.791	-0.737	1.050	0.159
0.002	0.0180	0.0443	0.0576	-0.347	0.895	-0.869	1.139	0.181

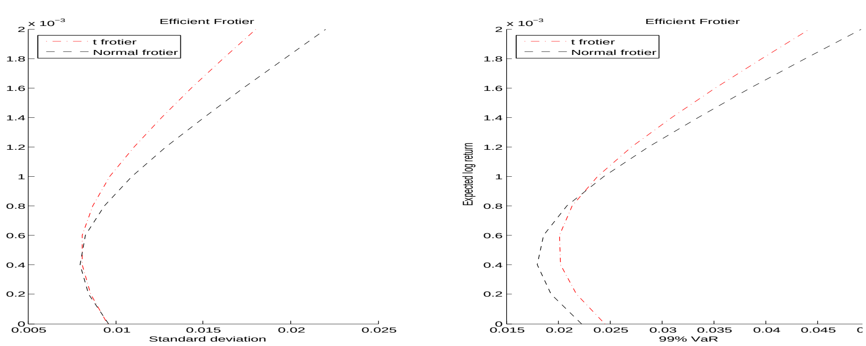

Figure 4.7: Student t and normal efficient frontier versus standard deviation and 99%VaR

distribution. If we use other measures such as VaR or standard deviation, we will get opposite conclusion as figure 4.7 shows that when the expected return is greater than some point, the normal frontier is contained in t frontier.

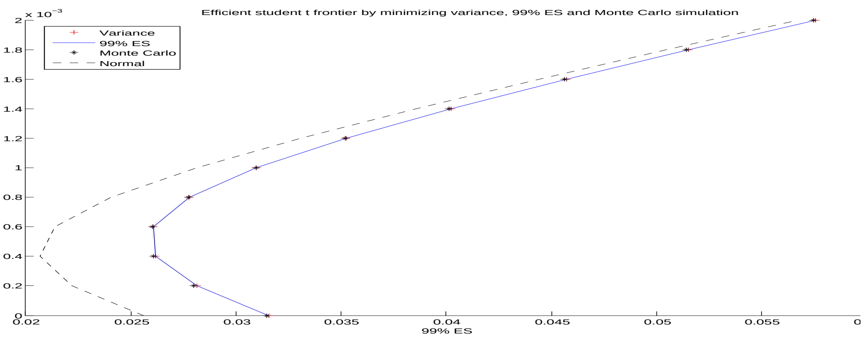

Figure 4.8: Student t efficient frontier and normal frontier versus 99% ES

Skewed t Frontier

Table 4.8 shows the location parameters and skewness parameters for the filtered data using skewed Student t distribution. The degree of freedom is 5.93. Table 4.9 shows the dispersion matrix(upper triangular), and correlation matrix(lower triangular) for the 5 stocks.

Table 4.8: Expected filtered log return and skewness parameters for skewed t distribution

Stock	Disney	Exxon	Pfizer	Altria	Intel
location parameters	-0.071	0.089	-0.030	0.161	0.042
skewness parameters	0.073	-0.010	0.010	-0.079	-0.010
$GARCH$ volatility	0.0107	0.0128	0.0130	0.0113	0.0156

We use Monte Carlo simulation[††] to get skewed t frontier by minimizing expected

[††]We use the filtered returns series to calibrate skewed t distribution and then use the mean-variance mixture definition to sample the multivariate skewed t distribution to get the 1,000,000 samples $\hat{X}_{1000000\times 5}$. Specifically, in Matlab, we generate 1,000,000 multivariate normal distributed random variables with mean $\mathbf{0}$ and covariance $\hat{\Sigma}$, which is calibrated using filtered returns series, then we generate 1,000,000 $InverseGamma(\nu/2, \nu/2)$ distributed random variables, finally, we get 1,000,000 multivariate skewed t

Table 4.9: Dispersion and correlation matrix of skewed t distribution: the diagonal are diagonal of dispersion matrix, the upper triangular is dispersion matrix and the lower triangular is correlation matrix

Stock	Disney	Exxon	Pfizer	Altria	Intel
Disney	0.706	0.269	0.267	0.164	0.333
Exxon	0.360	0.773	0.275	0.171	0.244
Pfizer	0.377	0.373	0.704	0.157	0.251
Altria	0.254	0.273	0.258	0.509	0.139
Exxon	0.458	0.324	0.349	0.226	0.736

shortfall. Table 4.10 shows the 99% level portfolio compositions and the corresponding 99% ES and table 4.11 shows the 95% level portfolio compositions and the corresponding 95% ES. From table 4.10, we can see that the portfolio compositions in skewed t frontier are different from those compositions in normal and Student t frontier at 99% level. Since skewed t distribution is not of elliptical distributions, the 99% level and 95% level produce slightly different portfolio compositions.

Table 4.10: Portfolio composition and corresponding 99%ES for skewed t frontier

Return	99% ES	Disney	Exxon	Pfizer	Altria	Intel
0	0.0320	0.393	-0.219	0.515	0.337	-0.026
0.0002	0.0280	0.383	-0.058	0.328	0.363	-0.015
0.0004	0.0263	0.374	0.101	0.139	0.389	-0.003
0.0006	0.0274	0.381	0.259	-0.051	0.406	0.006
0.0008	0.0312	0.399	0.415	-0.244	0.416	0.013
0.001	0.0367	0.428	0.573	-0.436	0.415	0.021
0.0012	0.0433	0.456	0.733	-0.626	0.413	0.024
0.0014	0.0506	0.485	0.892	-0.817	0.409	0.031
0.0016	0.0583	0.514	1.052	-1.007	0.406	0.036
0.0018	0.0662	0.549	1.209	-1.200	0.404	0.038
0.002	0.0744	0.587	1.365	-1.394	0.399	0.042

distributed random variables by using the mean-variance mixture definition. The restored samples $X = \hat{X}A$, where $A = Diag(\boldsymbol{\sigma})$. The restored mean $\mu_i = (\hat{\mu}_i + \frac{\nu}{\nu-2}\hat{\gamma}_i)\sigma_i$ where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\gamma}}$ are location and skewness parameters respectively calibrated using filtered data.

Table 4.11: Portfolio composition and corresponding 95%ES for skewed t frontier

Return	95% ES	Disney	Exxon	Pfizer	Altria	Intel
0	0.0215	0.354	-0.222	0.515	0.367	-0.013
0.0002	0.0187	0.348	-0.065	0.325	0.393	-0.002
0.0004	0.0175	0.349	0.094	0.136	0.415	0.006
0.0006	0.0182	0.356	0.253	-0.054	0.430	0.014
0.0008	0.0206	0.369	0.412	-0.245	0.442	0.023
0.001	0.0242	0.386	0.570	-0.435	0.449	0.029
0.0012	0.0285	0.407	0.730	-0.625	0.453	0.036
0.0014	0.0333	0.426	0.889	-0.816	0.459	0.042
0.0016	0.0383	0.447	1.047	-1.007	0.466	0.047
0.0018	0.0436	0.470	1.206	-1.197	0.468	0.053
0.002	0.0489	0.492	1.366	-1.387	0.472	0.057

From figure 4.9, we can see the t frontier can not be reached at both 95% and 99% level if we believe in skewed t distribution. Multivariate skewed t distribution fits financial data a little better than t distribution because of its ability of modeling the skewness.

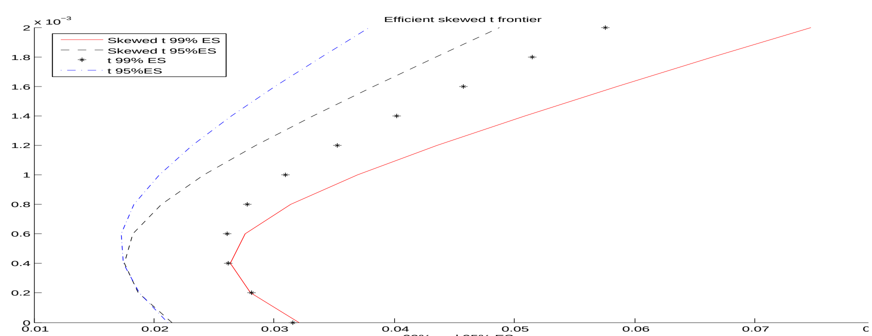

Figure 4.9: Skewed t efficient frontier at 99% ES or 95% ES versus t frontier

4.5 Hedge Equity Portfolio Using Index Futures

There are many index futures which are actively traded in CME and CBOT, especially the mini versions, such as the mini Dow and mini S&P500. Other actively traded futures are Midcap 400, Nasdaq 100 and Russell 2000.

The Dow has only 30 components so that it is not difficult to replicate this index. If we buy the portfolio, we want to sell a certain amount of mini Dow futures to hedge the position. We can get the hedge ratio by a simple linear regression of dollar value change in portfolio versus dollar value change in future or a simple ratio of the dollar value of portfolio to the face value of future. The transaction cost of futures is less than for equity and it is more liquid than equity so that it is easy to trade.

4.6 Conclusion

Standard deviation and VaR usually are not coherent risk measures, while ES is. Under an elliptical distribution, no matter what confidence level it is, minimizing any translation invariant and positive homogenous risk measure will produce the same portfolio composition as Markowitz style portfolio composition given a certain return. Only the difference in distribution leads to different portfolio compositions. We should choose heavy tailed distributions to model the portfolio risk.

We use multivariate generalized hyperbolic distributions to fit the filtered *i.i.d.* five dimensional returns series. Skewed t has the largest log likelihood and fewest number of parameters among the four tested generalized hyperbolic distributions. Student t has the second largest log likelihood. Student t is elliptical, while skewed t is not.

We try to get the normal, Student t and skewed t efficient frontier under risk measure ES. The portfolio composition is different at a given return if we choose different distributions. For normal frontier or Student t frontier, different confidence levels still lead to the same portfolio compositions, while for skewed t distribution, different confidence levels lead to slightly different portfolio compositions. Under 99% ES, if we believe in Student t distribution, normal frontier is unreachable, and if we believe in skewed t distribution, Student t frontier is unreachable.

CHAPTER 5

Portfolio Credit Risk

Copulas have become a popular way to describe and construct the dependence structure of multivariate distributions. There are two types of copulas. The first type of copulas are the bivariate Archimedean copulas, which have been successfully used in the modeling of bivariate distributions. We can follow Frees and Valdez (1997) to choose the best-fit Archimedean copula. The Archimedean copula can be extended to n dimensions, but we will lose almost all dependence information since all pairs of variables have the same Kendall's tau correlation coefficient. The second type of copulas are the elliptical copulas. Gaussian and Student t distributions are elliptical distributions. Gaussian and Student t copulas are derived from Gaussian and Student t distributions, respectively.

Corresponding to the heavy tail property in univariate distributions, tail dependence is used to model the co-occurrence of extreme events. Upper tail dependence is used to model the probability that two uniform random variables tend to go to the largest value 1 together; lower tail dependence similarly for the smallest value 0.

The success of copulas greatly depends on the availability of a fast algorithm to calculate the cumulative distribution functions (CDF) and quantiles of the corresponding one dimensional Gaussian and Student t distributions. The Student t copula is tail dependent while the Gaussian copula is tail independent. Schmidt(2003a) showed that all generalized hyperbolic copulas derived from symmetric generalized hyperbolic distributions are tail independent except for the Student t copula. The calibration of a Student t copula is very fast if we fix the degree of freedom ν, which in turn is optimized by maximizing a log likelihood. Detailed algorithms for calibrating Student t copulas can be found in the work of many researchers, such as Di Clemente and Romano (2003a), Demarta and McNeil (2005), Mashal and Naldi(2002) and Galiani(2003). The calibration of a Student t copula is separate from the calibration of marginal distributions. It is generally suggested to use the empirical distributions to fit the margins, but empirical distributions tend to have poor performance in the tails. A hybrid of the parametric and non parametric method considers the use of the empirical distribution in the body and generalized Pareto distribution (GPD) in the tails. Some use Gaussian distribution in the body. Di Clemente and Romano (2003a) used Student t copula and Gaussian distribution in the center and left tail and GPD in the right tail for the margins to model the multivariate losses. We will be able to avoid these issues because we can effectively calibrate the full distribution directly by using skewed t or Student t distributions.

Credit risk modeling has two popular approaches, structural and reduced form. In the structural approach, in which default is modeled as a first passage time across a barrier, the default time τ is predictable, so that the short credit spread tends to zero, which is not consistent with market observations. In reduced form, or stochastic intensity modeling, τ is unpredictable and the short credit spread tends to the intensity. In addition, the stochastic intensity can be modeled like an interest rate process. For these reasons we focus on the reduced form approach to credit risk models in this dissertation. An excellent book about interest rate modeling is Brigo and Mercurio(2005). Another is Schönbucher (2003), who uses the Heath-Jarrow-Morton framework to describe the term structure of a defaultable bond.

In the pricing of multiname credit derivatives such as basket credit default swaps (CDS), and collateralized debt obligations (CDO), the most important issue is the correlations between those default obligors. Defaults are rarely observed. Copulas can be introduced to model these correlations by using the correlations of corresponding equity prices. We show that Kendall's τ correlation remains invariant under monotone transformations. This is the foundation of modeling the correlation of credit events by using the correlation of underlying equities via copulas, though nobody mentions this correlation invariance property. Under our setting of portfolio credit risk, the default times are increasing functions of copula uniform random variables, which are the cumulative distribution function(CDF) transformations of corresponding equity prices, so that default times are increasing transformations of corresponding equity prices. Therefore, the Kendall's τ correlations of default times are the same as the correlations of corresponding equities. Recently, several researchers have discussed the pricing of basket CDS and CDO via copulas, such as Galiani (2003), Mashal and Naldi (2002), and Meneguzzo and Vecchiato (2002), among others.

In fact, we can also use a multivariate distribution of underlying equity prices instead of using copulas to model the correlation of credit events with the correlation of underlying equities by using this correlation invariance property. The only difference between a copula approach and a distribution approach lies in the calibration procedure. For a copula approach, the calibrations of marginal distributions and copula are separate, while for a distribution approach, the two are jointly calibrated.

There are many applications of multivariate distributions or multivariate copulas. Di Clemente and Romano (2003b) used copulas to minimize expected shortfall(ES) in modeling operational risk. Di Clemente and Romano (2004) applied the same framework in the portfolio optimization of CDS. Masala, Menzietti and Micocci (2004) used the Student t copula and the transition matrix with gamma distributed hazard rate and beta distributed recovery rate to get the efficient frontier for credit portfolios by minimizing ES.

We applied the generalized hyperbolic distributions to model the multivariate equity returns by using the EM algorithm in last chapter. We showed that skewed t has better performance and a fast algorithm compared with other generalized hyperbolic distributions. Furthermore, for the Student t distribution, we have greatly simplified formulas and a faster algorithm than for the skewed t distribution. For the Student t copula, there is still no good method to calibrate the degree of freedom ν except to find it by direct search. The calibration of student t copula takes days while the calibration of skewed t or Student t distribution takes minutes.

We use the Student t and skewed t distributions to model the correlations of default obligors in this final chapter.

This chapter is organized as following. In Section 5.1, we follow Rutkowski (1999) to price the single name credit risk. We give an introduction to copula in Section 5.2. We show that the rank correlation Kendall's tau remains invariant under monotone transformation in Section 5.3. In Section 5.4, we follow Schönbucher (2003) to provide our model setup for calculating default probabilities for the k-th to default using copulas. In addition, we show that we can use the distributions of underlying equities to model the correlations instead of using copulas. In Section 5.5, we do some default probabilities analysis for the k-th to default using different copulas. In Section 5.6, we use (skewed) t distributions and t copula to price the basket credit default swaps. To our knowledge, we are the first to apply skewed

t or Student t distribution to price basket credit default swaps.

5.1 An Introduction to Credit Risk

There are two popular approaches in the credit risk modeling: the reduced form approach and the structural approach. In structural modeling, the default time τ is predictable, so that the short credit spread tends to zero, which is not consistent with market observation. In reduced form modeling, τ is unpredictable and the short credit spread tends to the intensity. Many people call the reduced form modeling stochastic intensity modeling.

For the single name credit risk, we follow the reduced form approach of Rutkowski (1999). Suppose that τ is the default time of a firm. Let $H_t = I_{\tau \le t}$, and $\mathcal{H}_t = \sigma(H_s : s \le t)$ denote the default time information filtration. We denote by F the right-continuous cumulative distribution function of τ, i.e., $F(t) = P(\tau \le t)$.

Definition 5.1.1 Hazard Function. *The function* $\Gamma : \mathbb{R}^+ \to \mathbb{R}^+$ *given by*

$$\Gamma(t) = -\log(1 - F(t)), \forall t \in \mathbb{R}^+ \tag{5.1.1}$$

is called the hazard function or cumulative hazard function. If F is absolutely continuous, i.e., $F(t) = \int_0^t f(u)du$, then we have

$$F(t) = 1 - e^{-\int_0^t \lambda(u)du}, \tag{5.1.2}$$

where λ is called intensity function or hazard rate.

The following is the probability density function of default time τ

$$f(t) = \lambda(t)S(t), \tag{5.1.3}$$

where $S(t) = 1 - F(t)$.

Suppose the risk free interest rate r is a non-negative deterministic function so that the price at time t of a unit of default free zero coupon bond with maturity T equals $B(t,T) = e^{-\int_t^T r(u)du}$.

We have the following pricing formulas when the only information is default time.

Proposition 5.1.2 Rutkowski(1999). *Assume that $t \le T$. If*

$$Y_t = I_{\{t<\tau\le T\}} h(\tau) e^{-\int_t^\tau r(u)du} + I\{\tau > T\} c e^{-\int_t^T r(u)du},$$

where c is a constant, and if $\Gamma(t)$ is absolutely continuous, then

$$E(Y_t|\mathcal{H}_t) = I_{\{\tau>t\}} \left(\int_t^T h(u)\lambda(u) e^{-\int_t^u \hat{r}(v)dv} du + ce^{-\int_t^T \hat{r}(u)du} \right), \tag{5.1.4}$$

where $\hat{r}(v) = r(v) + \lambda(v)$.

The first term is the price of default payment, while the second term is the price of survival payment. Note that in the first term we use equation 5.1.3 to denote the probability density function of τ. The second term tells us that a defaultable instrument can be valued as if it were default free by replacing the interest rate by the sum of interest rate and default intensity. We use this proposition to price basket credit default swaps.

Let $\mathcal{F}_t = \sigma(V_s : s \leq t)$, where$\{V_t : 0 \leq t \leq T\}$ is the firm's price series, denote the asset price information, and $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$ denote all the information, until time t. Let us define $F_t = P(\tau \leq t|\mathcal{F}_t)$.

Definition 5.1.3 Martingale Hazard Process. *Assume that $F_t < 1$ for every $t \in \mathbb{R}^+$. The $\mathcal{F}$-*martingale hazard process *of τ denoted by Γ is defined through*

$$\exp(-\Gamma_t) = P(\tau > t|\mathcal{F}_t) = 1 - F_t. \tag{5.1.5}$$

We have the following pricing formulas when we have all the information.

Proposition 5.1.4 Rutkowski(1999). *Assume that $t \leq T$. If Γ is well defined, then we have*

$$E[I_{\{\tau>t\}}Y|\mathcal{G}_t] = I_{\{\tau>t\}} \frac{E[I_{\{\tau>t\}}Y|\mathcal{F}_t]}{P[\tau > t|\mathcal{F}_t]} = I_{\{\tau>t\}} E[I_{\{\tau>t\}} Y e^{\Gamma_t}|\mathcal{F}_t]. \tag{5.1.6}$$

Furthermore, if Y is a $\mathcal{F}_T$-measurable random variable.

$$E[I_{\{\tau>T\}}Y|\mathcal{G}_t] = I_{\{\tau>t\}} E[\exp(\Gamma_t - \Gamma_T)Y|\mathcal{F}_t]. \tag{5.1.7}$$

This proposition tells us the pricing under all information is reduced to the pricing under asset information filtration.

Corollary 5.1.5 *For $t \leq T$, we have*

$$P[\tau > T|\mathcal{G}_t] = I_{\{\tau>t\}} \frac{P[\tau > T|\mathcal{F}_t]}{P[\tau > t|\mathcal{F}_t]}. \tag{5.1.8}$$

Definition 5.1.6 *If Γ is absolutely continuous, i.e. $\Gamma_t = \int_0^t \lambda_u du$, then λ is called the $\mathcal{F}$-intensity of a random time τ.*

Corollary 5.1.7 Stochastic Intensity. *For $t \leq T$, we have the following so called* stochastic intensity*,*

$$\lambda_t I_{\{\tau>t\}} = -\frac{\partial}{\partial T} P[\tau > T|\mathcal{G}_t]|_{T=t} = \lim_{T\downarrow t} \frac{1}{T-t} P[t < \tau \leq T|\mathcal{G}_t]. \tag{5.1.9}$$

5.2 An Introduction to Copula

Copulas are used to model dependence. One of the definitions can be found in Li (1999), the first one who uses copula to price portfolio credit risk.

5.2.1 Definition and Basic Properties of Copula Functions

Definition 5.2.1 Copula Function. *For d uniform random variables,$U_1, U_2, \cdots, U_d$, the joint distribution function C,defined as*

$$C(u_1, u_2, \cdots, u_d) = P[U_1 \leq u_1, U_2 \leq u_2, \cdots, U_d \leq u_d],$$

is called a copula *function.*

Proposition 5.2.2 *Sklar's Theorem(Sklar(1959)).* *Let F be a joint distribution function with margins $F_1, F_2, \cdots, F_d$, then there exists a copula C such that for all $(x_1, x_2, \cdots, x_d) \in \mathbb{R}^d$,*

$$F(x_1, x_2, \cdots, x_d) = C(F_1(x_1), F_2(x_2), \cdots, F_d(x_d)). \tag{5.2.1}$$

If $F_1, F_2, \cdots, F_d$ are continuous, then C is unique. Conversely, if C is a copula and $F_1, F_2, \cdots, F_d$ are distribution functions, then the function F defined by equation 5.2.1 is a joint distribution function with margins $F_1, F_2, \cdots, F_d$.

The name copula means a function that couples a joint distribution function to its marginal distributions. If $X_1, X_2, \cdots, X_d$ are random variables with distributions $F_1, F_2, \cdots, F_d$ respectively and a joint distribution F, then copula C is also called the copula of $X_1, X_2, \cdots, X_d$, and $(U_1, U_2, \cdots, U_d) = (F_1(X_1), F_2(X_2), \cdots, F_d(X_d))$ also has a copula C. We will use this property to price basket credit default swaps later.

We usually assume the marginal distributions to be empirical distributions. Suppose that the sample data is $\mathbf{x_i} = (x_{i,1}, \cdots, x_{i,d})$, where $i = 1, \cdots, n$, then the empirical estimator of jth marginal distribution function

$$\hat{F}_j(x) = \frac{\sum_{i=1}^{n} I_{\{x_{i,j} \leq x\}}}{n+1}. \tag{5.2.2}$$

Demarta and McNeil(2005) suggested to divide by $n+1$ to keep the estimation away from the boundary 1. By using different copulas and empirical or other margins, we can create a rich family of multivariate distributions. It is not required that the margins and joint distribution must be the same type of distribution; however, when we make copulas, we do require this, which can be seen from following corollary.

Corollary 5.2.3 *If $F_1, F_2, \cdots, F_m$ are continuous, then, for any $\mathbf{u} = (u_1, \cdots, u_m) \in [0,1]^m$, we have*

$$C(u_1, u_2, \cdots, u_m) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \cdots, F_m^{-1}(u_m)), \tag{5.2.3}$$

where $F_i^{-1}(u_i)$ denotes the inverse of the cumulative distribution function, namely, for $u_i \in [0,1]$, $F_i^{-1}(u_i) = \inf\{x : F_i(x) \geq u_i\}$.

Two types of copulas are widely used now. They are elliptical copulas and Archimedean copulas. Elliptical copulas are created from multivariate elliptical distributions. Gaussian and Student t are elliptical distributions. Gaussian and t copula are multivariate elliptical copulas which are created from Gaussian and Student t distribution respectively. Frank, Gumbel and Clayton copula are bivariate Archimedean copulas.

The copula remains invariant under increasing transformation (See Section 5.3). The copula remains invariant under the standardizations of marginal distributions. This means we can create elliptical copulas from standardized multivariate elliptical distributions.

Definition 5.2.4 Multivariate Gaussian Copula. *Let R be a symmetric, positive definite matrix with diag$(R) = \mathbf{1}$ and let Φ_R be the standardized multivariate normal distribution function with correlation matrix R. Then the* multivariate Gaussian copula *is defined as*

$$C(u_1, u_2, \cdots, u_m; R) = \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \cdots, \Phi^{-1}(u_m)), \tag{5.2.4}$$

where $\Phi^{-1}(u)$ denotes the inverse of the standard univariate normal cumulative distribution function.

Definition 5.2.5 Multivariate Student's t Copula. *Let R be a symmetric, positive definite matrix with $diag(R) = \mathbf{1}$ and let $T_{R,\nu}$ be the standardized multivariate Student t distribution function with correlation matrix R and ν degrees of freedom. Then the* multivariate Student t copula *is defined as*

$$C(u_1, u_2, \cdots, u_m; R, \nu) = T_{R,\nu}(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2), \cdots, T_\nu^{-1}(u_m)), \tag{5.2.5}$$

where $T_\nu^{-1}(u)$ denotes the inverse of standard univariate Student t cumulative distribution function.

Definition 5.2.6 Archimedean Copula. *An* Archimedean Copula function $C : [0,1]^2 \to [0,1]$ *is an function which can be represented in the following form:*

$$C(u,v) = \phi^{[-1]}\left(\phi(u) + \phi(v)\right), \tag{5.2.6}$$

where $\phi : [0,1] \to [0,\infty]$ is a continuous, strictly decreasing and convex function satisfying $\phi(1) = 0$, and ϕ is called the generator of the copula. The pseudo-inverse of ϕ , $\phi^{[-1]}$: $[0,\infty] \to [0,1]$ is given by

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & \textit{for } 0 \le t \le \phi(0) \\ 0 & \textit{for } \phi(0) \le t \le \infty. \end{cases} \tag{5.2.7}$$

Furthermore, if $\phi(0) = \infty$, then $\phi^{[-1]} = \phi^{-1}(t)$, and we call ϕ and C respectively a strict generator and a strict Archimedean copula.

Gumbel Copula. Let $\phi(t) = (-\ln t)^\theta$ with $\theta \ge 1$. Then, from equation(5.2.6), we have

$$C_\theta^{Gumbel}(u,v) = \phi^{-1}[\phi(u) + \phi(v)] = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right\}. \tag{5.2.8}$$

Clayton Copula. Let $\phi(t) = (t^{-\theta} - 1)/\theta$ with $\theta \in [-1,\infty)\backslash\{0\}$. Then, from equation(5.2.6), we have

$$C_\theta^{Clayton}(u,v) = \max\left[(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0\right]. \tag{5.2.9}$$

Note that if $\theta > 0$, then $\phi(0) = \infty$, and we can get the strict copula,

$$C_\theta^{Clayton}(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}. \tag{5.2.10}$$

Frank Copula. Let $\phi(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$ with $\theta \in \mathbb{R}\backslash\{0\}$. Then, from equation(5.2.6), we have

$$C_\theta^{Frank}(u,v) = -\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right\}. \tag{5.2.11}$$

A lot of other copula generators can be found in Joe(1997). The Archimedean copulas can be used to construct a very rich family of bivariate distributions. Frees and Valdez(1997) provides a way to choose best-fit bivariate Archimedean copulas. Archimedean copulas have only one parameter so that they are easy to calibrate.

5.3 Measure of Dependence

In this chapter, when we talk about correlation, we mean Kendall's tau rank correlation.

5.3.1 Rank Correlation

Definition 5.3.1 Kendall's Tau. *Kendall's tau correlation for the bivariate random vector* (X, Y) *is defined as*

$$\tau(X,Y) = P((X-\hat{X})(Y-\hat{Y}) > 0) - P((X-\hat{X})(Y-\hat{Y}) < 0), \tag{5.3.1}$$

where $(\hat{X}, \hat{Y})$ *is an independent copy of* (X, Y).

As suggested by Meneguzzo& Vecchiato(2002), the sample consistent estimator of Kendall's tau is given by

$$\hat{\tau} = \frac{\sum_{i,j=1,i<j}^{n} sign[(x_i - x_j)(y_i - y_j)]}{n(n-1)/2}, \tag{5.3.2}$$

where $sign(x) = 1$ if $x \geq 0$, otherwise $sign(x) = 0$, and n is the number of observations.

Nelson(1999) showed that Kendall's tau τ depends only on the copula of (X, Y), and it is given by

$$\tau(X,Y) = 4\int\int_{[0,1]^2} C(u,v)dC(u,v) - 1. \tag{5.3.3}$$

It has nothing to do with the marginal distributions. Sometimes, we may need the following formula

$$\tau(X,Y) = 1 - 4\int\int_{[0,1]^2} C_u(u,v)C_v(u,v)dudv, \tag{5.3.4}$$

where C_u denotes the partial derivative of $C(u, v)$ with respect to u and C_v denotes the partial derivative of $C(u, v)$ with respect to v.

Lindskog et al. (2003) showed that for elliptical distributions,

$$\tau(X,Y) = \frac{2}{\pi}\arcsin(\rho), \tag{5.3.5}$$

where ρ is the Pearson's linear correlation coefficient between random variable X and Y.

Nelson (1999) showed that for Archimedean copulas,

$$\tau(X,Y) = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt. \tag{5.3.6}$$

We can use equation 5.3.2 to get the sample Kendall's tau and then use equation 5.3.5 or equation 5.3.6 to estimate the copula parameters. This is a very rough solution. Most people use this method to get an initial guess of copula parameters when they calibrate a copula.

Proposition 5.3.2 Copula of Transformations(Nelson 1999). *Let X and Y be continuous random variables with copula C_{XY}. If both $\alpha(X)$ and $\beta(Y)$ are strictly increasing on RanX and RanY respectively, then $C_{\alpha(X)\beta(Y)} = C_{XY}$. If both $\alpha(X)$ and $\beta(Y)$ are strictly decreasing on RanX and RanY respectively, then $C_{\alpha(X)\beta(Y)}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v)$.*

Theorem 5.3.3 Invariability of Kendall's Tau Under Monotone Transformation. *Let X and Y be continuous random variables with copula C_{XY}. If both $\alpha(X)$ and $\beta(Y)$ are strictly increasing or strictly decreasing on RanX and RanY respectively, then $\tau_{\alpha(X)\beta(Y)} = \tau_{XY}$.*

Proof: we just need to show the second part. If both $\alpha(X)$ and $\beta(Y)$ are strictly decreasing, then $C_{\alpha(X)\beta(Y)}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v)$. From equation 5.3.4, we have

$$\tau_{\alpha(X)\beta(Y)} = 1 - 4\int\int_{[0,1]^2}(1 - C_1(1-u,1-v))(1 - C_2(1-u,1-v))dudv$$

where C_i denotes the partial derivative with respect to ith variable to avoid confusion. By replacing $1-u$ by x and $1-v$ by y, we have

$$\tau_{\alpha(X)\beta(Y)} = 1 - 4\int\int_{[0,1]^2}(1 - C_1(x,y))(1 - C_2(x,y))dxdy.$$

Since

$$\int\int_{[0,1]^2} C_1(x,y)dxdy = \int_{[0,1]} ydy = 0.5,$$

we have $\tau_{\alpha(X)\beta(Y)} = \tau_{XY}$. ■

This theorem is the foundation of modeling of default correlation in the pricing of portfolio credit risk. In fact, the default time is a increasing transformation of the price of underlying equity, which can seen in Section 5.4, so that the correlation of default times is the same as the correlation of underlying equities. Most people use a copula to model the correlation by assuming the marginal distributions. No-one seems to be aware, that, actually, we can use the distribution of equity prices to model the correlation of default times too. We do not need to assume the marginal distributions at all when we model the correlations via the multivariate distribution. The calibration of multivariate distributions is integrative, while calibration via copulas is separated into two steps, where one is for marginal distributions and another one is for dependence.

5.3.2 Tail Dependence

Corresponding to the heavy tail property in univariate distributions, tail dependence is used to model the co-occurrence of extreme events. For credit risk, firms tend to crash together because of their connection in business.

Definition 5.3.4 Tail Dependence Coefficient(TDC). *Let* (X_1, X_2) *be a bivariate vector of continuous random variables with marginal distribution function* F_1 *and* F_2*. The level of* upper tail dependence λ_U *and* lower tail dependence λ_L *are given respectively by,*

$$\lambda_U = \lim_{u \uparrow 1} P[X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)], \tag{5.3.7}$$

$$\lambda_L = \lim_{u \downarrow 0} P[X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)]. \tag{5.3.8}$$

If $\lambda_U \in (0, 1]$*, then the two random variables* (X_1, X_2) *are said to be asymptotically dependent in the upper tail. If* $\lambda_U = 0$*, then the two random variables* (X_1, X_2) *are said to be* asymptotically independent in the upper tail. *If* $\lambda_L \in (0, 1]$*, then the two random variables* (X_1, X_2) *are said to be asymptotically dependent in the lower tail. If* $\lambda_L = 0$*, then the two random variables* (X_1, X_2) *are said to be* asymptotically independent in the lower tail.

Joe(1997) gave the copula version of TDC,

$$\lambda_U = \lim_{u \uparrow 1} \frac{[1 - 2u + C(u, u)]}{1 - u}, \tag{5.3.9}$$

$$\lambda_L = \lim_{u \downarrow 0} \frac{C(u, u)}{u}. \tag{5.3.10}$$

For elliptical copulas, $\lambda_U = \lambda_L$, and we just denote them by λ. Embrechts et al. (2001) showed that for a Gaussian copula, $\lambda = 0$, and for a Student t copula,

$$\lambda = 2 - 2t_{\nu+1}\left(\sqrt{\nu+1}\frac{\sqrt{1-r}}{\sqrt{1+r}}\right). \tag{5.3.11}$$

We can see that λ is an increasing function of the Pearson correlation coefficient r and a decreasing function of the degree of freedom ν. The Student t copula is a tail dependent copula. We can see the difference of the tail dependence between Gaussian copula and Student t copula from figure 5.1.

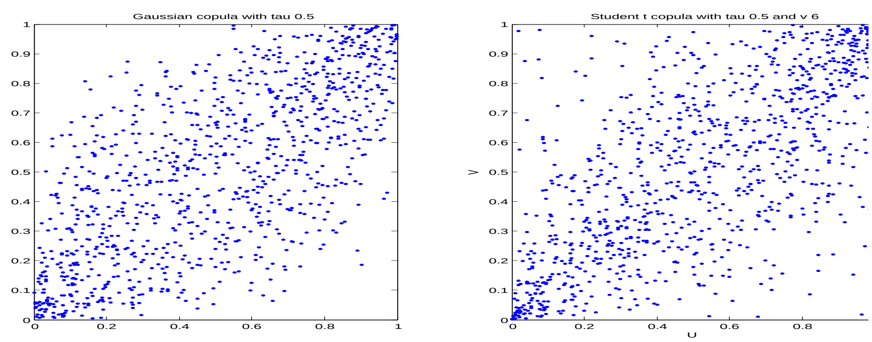

Figure 5.1: 1000 samples of Gaussian and Student t copula with Kendall's $\tau = 0.5$. There are more points in both corners for Student t copula.

The Clayton copula is lower tail dependent for $\theta > 0$ and $\lambda_L = 2^{-1/\theta}$, but it is upper tail independent.

The Gumbel copula is upper tail dependent for $\theta > 1$ and $\lambda_U = 2 - 2^{1/\theta}$, but it is lower tail independent.

The Frank copula is tail independent. We can see difference in the tail dependence among these three Archimedean copulas from figure 5.2.

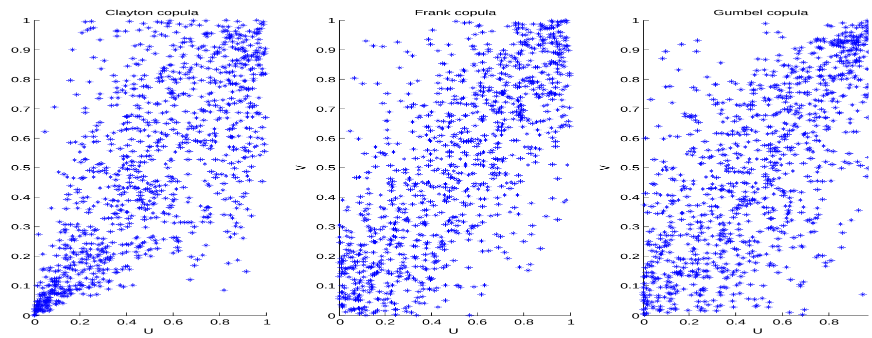

Figure 5.2: 1000 samples of Clayton, Frank, and Gumbel copula with Kendall's $\tau = 0.5$

5.4 Portfolio Credit Risk

5.4.1 Copula Approach Setup

We provide our setup of portfolio credit risk by following Schönbucher(2003). Suppose that we are standing at time t=0.

Assumption 5.4.1 *Suppose there are d firms. For each obligor $i \leq d$, we define*

1. The default intensity $\lambda^i(t)$: a deterministic function. We usually assume it to be a step function.

2. The survival function $S^i(t)$:

$$S^i(t) := \exp\left(-\int_0^t \lambda^i(u)du\right). \tag{5.4.1}$$

3. The default trigger variables $1 - U_i$: uniform random variables on $[0, 1]$. The d-dimensional vector $\mathbf{1} - \mathbf{U} = (1 - U_1, 1 - U_2, \cdots, 1 - U_d)$ is distributed according to the d-dimensional copula C(cf. definition 5.2.1).

4. The time of default τ_i of obligor i, where $i = 1, \cdots, d$,

$$\tau_i := \inf\{t : S^i(t) \le 1 - U_i\}. \tag{5.4.2}$$

The copula of $\mathbf{U}$ is denoted by $\hat{C}$. It is also called survival copula of $\mathbf{1} - \mathbf{U}$. See Georges, Lamy, Nicolas, Quibel and Roncalli(2001) for more details about survival copulas.

From equation 5.4.2, we can see that the default time τ_i is a increasing transformation of uniform random variable U_i, so that the rank correlation Kendall's tau between default times is the same as Kendall's tau between the uniform random variables, and the copula of $\boldsymbol{\tau}$ equals the copula of $\mathbf{U}$, in other words, the copula of $\mathbf{1} - \mathbf{U}$, is the survival copula of $\boldsymbol{\tau}$. In fact, the defaults rarely happen. In practice, people usually take the correlation of underlying equities as the correlation of default times. The uniform random variables U_i satisfies $U_i = F_i(X_i)$, where F_i is the distribution function of obligor i's equity price X_i and we usually assume it to be empirical distribution. Therefore, τ_i is a increasing transformation of underlying equity price X_i so that the Kendall's tau between default times is the same as Kendall's tau between the underlying equities. The copula of $\mathbf{X}$ is the same as the copula of $\mathbf{U}$.

Theorem 5.4.2 Joint Survival Probabilities. *The joint survival probabilities of* $(\tau_1, \tau_2, \cdots, \tau_d)$ *are given by*

$$P[\tau_1 > T_1, \tau_2 > T_2, \cdots, \tau_d > T_d] = C(S^1(T_1), \cdots, S^d(T_d)). \tag{5.4.3}$$

Proof: From the definition of default in equation 5.4.2, we have

$$P[\tau_1 > T_1, \cdots, \tau_d > T_d] = P[1 - U_1 < S^1(T_1), \cdots, 1 - U_d < S^d(T_d)].$$

By the definition of copula of $\mathbf{1} - \mathbf{U}$ we have

$$P[\tau_1 > T_1, \cdots, \tau_d > T_d] = C\left(S^1(T_1), \cdots, S^d(T_d)\right),$$

where C is the copula of $\mathbf{1} - \boldsymbol{U}$ and it is also the survival copula of $(\tau_1, \tau_2, \cdots, \tau_d)$. ∎

Let us denote the default function, $F^i(t) = 1 - S^i(t)$.

Theorem 5.4.3 Joint Default Probabilities. *The joint default probabilities of* $(\tau_1, \tau_2, \cdots, \tau_d)$ *are given by*

$$P[\tau_1 \le T_1, \tau_2 \le T_2, \cdots, \tau_d \le T_d] = \hat{C}(F^1(T_1), \cdots, F^d(T_d)). \tag{5.4.4}$$

Proof: From the definition of default in equation 5.4.2, we have

$$P[\tau_1 \leq T_1, \cdots, \tau_d \leq T_d] = P[1 - U_1 \geq S^1(T_1), \cdots, 1 - U_d \geq S^d(T_d)].$$

By the definition of copula of $\mathbf{U}$ we have

$$P[\tau_1 \leq T_1, \cdots, \tau_d \leq T_d] = \hat{C}\left(F^1(T_1), \cdots, F^d(T_d)\right),$$

where $\hat{C}$ is the copula of $\boldsymbol{U}$ and it is also the copula of $(\tau_1, \tau_2, \cdots, \tau_d)$. ∎

In the Monte Carlo simulation, we calibrate and simulate the copula of $\mathbf{X}$, which is also the copula of $\boldsymbol{\tau}$ or $\mathbf{U}$, i.e. copula $\hat{C}$.

5.4.2 Distribution Approach

We can also use multivariate distributions of equity prices to model the correlation of default times. After we calibrate the full multivariate distribution, we can get the margins and take marginal transformations to get uniform random variables, $U_i = F_i(X_i)$, so that default time τ_i is a increasing transformation of underlying equity price X_i. In this way, the rank correlation Kendall's tau remains unchanged. As noted in Proposition 5.2.2, the copula of $\mathbf{U}$ is the same as the copula of $\mathbf{X}$. All the setup are the same as in the copula approach except the calibration procedure. The calibration of a Student t copula is separate from the calibration of marginal distributions. It is generally suggested to use the empirical distribution to fit the margins, but empirical distributions tend to have poor performance in the tails. A hybrid of the parametric and non parametric method considers the use of the empirical distribution in the body and generalized Pareto distribution (GPD) in the tails. Some use Gaussian distribution in the body. In addition, there is no good method to optimize the degree of freedom for Student t copula. We will be able to avoid these issues because we can effectively calibrate the full distribution such as Student t or skewed t directly.

In our setting, the default times are increasing transformations of corresponding equity prices. If the equity price is higher, the firm is healthier so that the default time is longer. It is noted that under the setting of Schönbucher(2003) of portfolio credit risk, the default times are decreasing transformations of corresponding equity prices. If we use symmetric copulas or distributions such as Student t copula, Gaussian copula and Student t distribution etc, then both settings are equivalent since for a symmetric copula C, the survival copula is itself. If we use skewed t distribution, then these two settings are different.

5.5 K-th to Default Probabilities Analysis Using Copulas

In this Section, we show by some experiments how to calculate the k-th to default probabilities, especially, first to default(FTD) and last to default (LTD) and sensitivity analysis of copula. K denotes the order of default in a basket of firms.

5.5.1 Algorithm

We calculate the k-th to default probabilities using the following procedures.

1. We use Matlab copula toolbox 1.0 to simulate Gaussian, Student t, Clayton and Gumbel copulas uniform variables $u_{i,j}$with the same Kendall's tau correlation, where $i = 1, \cdots, d$, $j = 1, \cdots, n$ and n is the number of samples *.

2. From equation(5.4.2), we get $\tau_{i,j}$ and sort according to column. The k-th row is a series of k-th to default times τ_i^k.

3. Divide the interval from year 0 to year 5 into 500 small sub intervals and count the number of points that τ_i^k falls into those sub intervals and divide by the number of samples to get the default probabilities during those small sub intervals. Finally, we get the default probability before time t by summing all the default probabilities over all the sub intervals before time t, where t is from 0 to 5.

In the following, we assume that there are two firms, there is no default until now, the Kendall's $\tau = 0.5$ for all copulas, the default intensity $\lambda_1 = 0.05$, $\lambda_2 = 0.03$, degree of freedom of t copula is 5, and all the Monte Carlo simulations have 1,000,000 runs.

5.5.2 The Default Probabilities of Last to Default(LTD)

We can see from Figure 5.3 that a copula function with lower tail dependence (Clayton copula) leads to highest default probabilities for LTD, while the copula function with upper

*Note that we use the copula of $\mathbf{U}$, $\hat{C}$, to simulate the copula random variables. This method is also used in the pricing of basket credit default swaps in the next section, where we calibrate the copula $\hat{C}$ from the underlying equity prices.

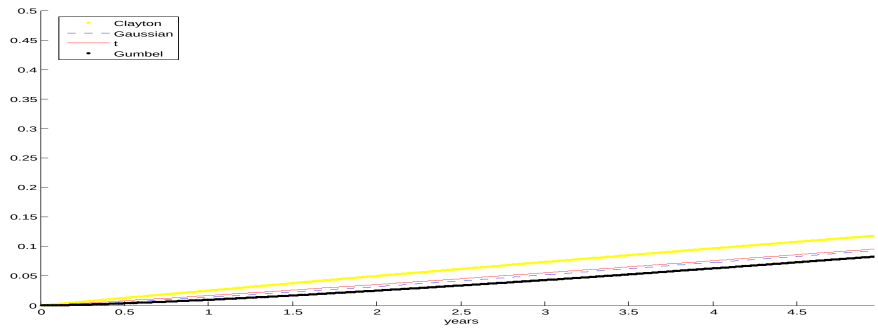

Figure 5.3: Default probabilities of LTD.

tail dependence (Gumbel copula) leads to lowest default probabilities. The tail dependent copula (Student t copula) leads to higher default probabilities than tail independent copula(Gaussian copula). The default events happen when the uniform random variables are small(close to 0). The last to default requires that both uniform variables in the basket are small. A lower tail dependent copula has more chance that both uniform variables are small than a tail independent copula.

5.5.3 The Default Probabilities of First to Default(FTD)

We can see from Figure 5.4 that a copula function with lower tail dependence (Clayton copula) leads to the lowest default probabilities for FTD, while a copula function with upper tail dependence(Gumbel copula) leads to the highest default probabilities. The first to default requires that only one or more of the uniform variables in the basket is small. Lower tail dependent copula has less chance that one or more uniform variables are small than tail independent copula.

Since defaults rarely occur, we are especially interested in the FTD. Archimedean copulas can only model bivariate distributions or correlations. The Student t copula is more promising than the Gaussian copula for its lower tail dependence.

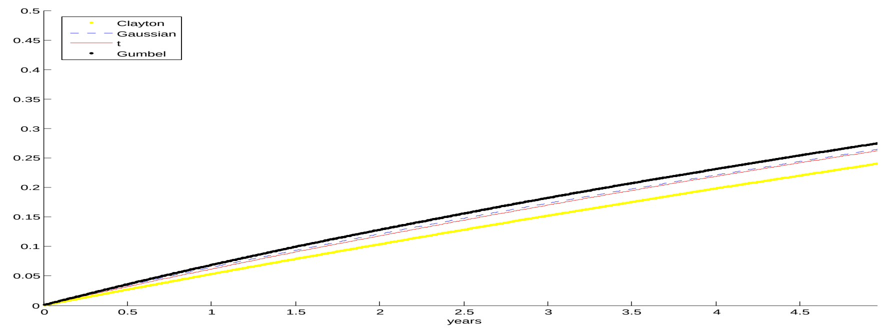

Figure 5.4: Default probabilities of FTD.

5.6 Pricing of Basket Credit Default Swap Using Elliptical Copulas and Skewed t Distribution

5.6.1 Credit Default Swaps

A credit default swap(CDS) is a contract that provides insurance against the risk of default of a particular company. The buyer of a CDS contract obtains the right to sell a particular bond issued by the company for its par value once a default occurs. The buyer pays a periodic payment, at time $t_1, \cdots, t_n$, as the fraction of nominal value M , to the seller until the end of the life of the contract $T = t_n$ or until a default time $\tau < T$ occurs. If a default occurs, the buyer still needs to pay the accrued payment from last last payment time to default time. There are $1/\Delta$ payments a year, and every payment is ΔkM. Usually, $\Delta = 1/2$.

5.6.2 Valuation of Credit Default Swaps

Set the current time $t_0 = 0$. Let us suppose the only information available is the default information, the interest rate follows a deterministic function, recovery rate is a constant and the expectation operator $E(\cdot)$ is under the risk neutral world. We use Proposition 5.1.2 to get the premium leg, accrued payment and default leg. The premium leg is the current price of periodic payments and the accrued payment is the current price of accumulated payment from last payment to default time. Default leg is the current price of default payment.

$$\begin{aligned} PL &= M\Delta k \sum_{i=1}^{n} E(B(0,t_i) I\{\tau > t_i\}) \\ &= M\Delta k \sum_{i=1}^{n} B(0,t_i) e^{-\int_0^{t_i} \lambda(u)du}, \end{aligned} \tag{5.6.1}$$

$$\begin{aligned} AP &= M\Delta k \sum_{i=1}^{n} E\left(\frac{\tau - t_{i-1}}{t_i - t_{i-1}} B(0,\tau) I\{t_{i-1} < \tau \le t_i\}\right) \\ &= M\Delta k \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \frac{u - t_{i-1}}{t_i - t_{i-1}} B(0,u)\lambda(u) e^{-\int_0^{u} \lambda(s)ds} du, \end{aligned} \tag{5.6.2}$$

$$\begin{aligned} DL &= M(1-R) E\left(B(0,\tau) I_{\{\tau \le T\}}\right) \\ &= M(1-R) \int_0^T B(0,u)\lambda(u) e^{-\int_0^{u} \lambda(s)ds} du. \end{aligned} \tag{5.6.3}$$

The spread price k^* is the k such that the value of credit default swap is zero,

$$PL(k^*) + AP(k^*) = DL(k^*). \tag{5.6.4}$$

5.6.3 Calibration of Default Intensity

Hull(2002) mentioned that the credit default swap market is so liquid that we can use the credit default swap spread data to calibrate the default intensity using equation 5.6.4.

In table 5.1, we have the credit default spread data on 07/02/2004 from *GFI* (http://www.gfigroup.com). The spread price is quoted in basis points. It is the payment made by the buyer of the *CDS* per year per dollar. The mid price is the average of bid price and ask price.

We denote the maturities of the *CDS* contract as $(T_1, \cdots, T_5) = (1,2,3,4,5)$. It is usually assumed that the default intensity is a step function, with step size to be 1 year, and it can be expressed in the following form,

$$\lambda(t) = \sum_{i=1}^{5} c_i I_{(T_{i-1},T_i]}(t)^{\dagger} \tag{5.6.5}$$

† $T_0 = 0$.

Table 5.1: Credit default swaps middle point quote

Company	Year 1	Year 2	Year 3	Year 4	year 5
AT&T	144	144	208	272	330
Bell South	12	18	24	33	43
Century Tel	59	76	92	108	136
SBC	15	23	31	39	47.5
Sprint	57	61	66	83	100

We can get c_1 by using the 1 year CDS spread price first. Knowing c_1, we can estimate c_2 using the 2 year CDS spread price. Following this procedure, we can estimate all the constant c_i for the default intensity.

In our calibration, we assume that recover rate is 0.4, risk free interest rate is 0.045 and the payment is paid semiannually. Under this setting, we can get PL, AP and DL explicitly. The calibrated default intensity is shown in table 5.2.

Table 5.2: Calibrated default intensity

Company	Year 1	Year 2	Year 3	Year 4	year 5
AT&T	0.0237	0.0237	0.0599	0.0893	0.1198
Bell South	0.0020	0.0040	0.0061	0.0105	0.0149
Century Tel	0.0097	0.0155	0.0210	0.0271	0.0469
SBC	0.0025	0.0052	0.0080	0.0109	0.0144
Sprint	0.0094	0.0108	0.0127	0.0235	0.0304

5.6.4 Pricing of Basket Credit Default Swaps

Suppose that there are I firms now. We try to price the 5 year basket credit default swaps. All the settings are the same as the single CDS except that the default event is triggered by the k-th default in the basket, where k is the seniority level of this structure. The seller of the basket CDS will face the default payment upon the k-th default, and the buyer will pay the spread price until k-th default or until maturity T. Let $(\tau_1, \cdots, \tau_I)$ denote the default order.

We can use Proposition 5.1.2 to get the premium leg, accrued payment and default leg,

$$PL = M\Delta k \sum_{i=1}^{n} E(B(0,t_i)I\{\tau_k > t_i\}), \tag{5.6.6}$$

$$AP = M\Delta k \sum_{i=1}^{n} E\left(\frac{\tau_k - t_{i-1}}{t_i - t_{i-1}} B(0,\tau_k) I\{t_{i-1} < \tau_k \leq, t_i\}\right), \tag{5.6.7}$$

$$DL = M(1-R)E\left(B(0,\tau)I_{\{\tau_k \leq T\}}\right). \tag{5.6.8}$$

The spread price k^* is the k such that the value of credit default swap is zero, i.e.,

$$PL(k^*) + AP(k^*) = DL(k^*). \tag{5.6.9}$$

The most important issue in the pricing of a basket CDS is the correlation of those defaults. Default data is rarely observable. It is a common experience that we use correlation of underlying equities to model the correlation of defaults because of the correlation invariance property though this theorem appears not to be stated in the literature.

It is not easy to get the distribution of τ_k so that it is difficult to calculate the expectations in all the legs. Monte Carlo simulation[‡] can be used to get the spread price k^* of k-th to default.

We use two approaches to apply the Monte Carlo simulation. One is the popular Student t copula approach. The Student t copula has tail dependence, which is good to model the co-occurance of credit events. However, it is also often argued that its symmetry and bivariate exchangeability are not realistic. We can also use skewed t distribution to model the correlation. The only difference between a distribution and a copula is that we need to specify marginal distributions for the copula, but we do not need to for the distribution approach. The calibration of a Student t copula is fast if we fix ν, but there is no good method to calibrate the degree of freedom ν. The calibration of a Student t copula for fixed ν can be found in a lot of current literature, such as Di Clemente and Romano (2003a), Demarta and McNeil (2005), and Galiani (2003). The skewed t distribution has better performance than the t distribution because of its ability to model skewness. It has heavier tails than the t distribution, which is good for risk management. At this time, we can not calibrate the skewed t copula suggested by Demarta and McNeil (2005), but we can try the skewed

[‡]Standard procedure can be found in Galiani(2003) etc.

t distribution since it is easy to calibrate and simulate by using the mean-variance mixture definition of GH discussed earlier.

We get the adjusted close prices for the five underlying stocks from finance.yahoo.com. The data is ranged from 07/02/1998 to 07/02/2004.

Copula Approach. We use the empirical distribution to model the marginal distributions and transform the equity prices into uniform variates. Then, we can estimate a Student t copula or Gaussian copula using those uniform variables. The log likelihood of the Gaussian copula is 936.90, while the log likelihood of the Student t copula is 1043.94. The Student t copula is far better than the Gaussian copula. The degree of freedom of the Student t copula is 7.406, which is found by maximizing log likelihood using direct search. The usual method is to loop ν from 2.001 to 20 and step size is 0.001. Each loop takes about 5 seconds and the calibration takes about 24 hours.

After we calibrate the copula, we sample the copula to get uniform random variables using the approach mentioned in Section 5.5, use equation 5.4.2 to get default times, and use equation 5.6.9 to find the spread price for k-th to default.

Distribution Approach. We calibrate the multivariate Student t or multivariate skewed t distribution first using EM algorithm discussed in chapter 2. The calibration is very fast, which takes less than 1 minute. The calibrated degree of freedom for both Student t and skewed t is 4.313. The log likelihood for skewed t is 18420.58, while for Student t is 18420.20. Then, we sample Student t or skewed t random variables and transform them into uniform random variables by the marginal CDF transformations. Then we follow the same procedure as for the copula approach to get the spread price for k-th to default.

Table 5.3: Spread price for k-th to default using different models

Model	FTD	2TD	3TD	4TD	LTD
Gaussian copula	525.6	141.7	40.4	10.9	2.2
Student t copula	506.1	143.2	46.9	15.1	3.9
Student t distribution	498.4	143.2	48.7	16.8	4.5
Skewed t distribution	499.5	143.9	49.3	16.8	4.5

We calculate the spread prices from FTD to LTD and report the results in table 5.3. We can see that lower tail dependent copula leads to higher default probability for LTD

and lower probability for FTD, thus leads to higher spread price for LTD and lower spread price for FTD. Student t has almost the same log likelihood and almost the same spread price of k-th to default as skewed t distribution. Both distributions lead to higher spread price for LTD and lower spread price for FTD.

The calibration of Student t distribution is integrated while to calibrate Student t copula, we need to assume marginal distributions first and we have no good method the calibrate the degree of freedom ν. Basket credit default swaps or collateralized debt obligations usually have a large number of securities. For example, a synthetic CDO called EuroStoxx50 issued on May 18, 2001 has 50 single name credit default swaps on 50 credits that belong to the DJ EuroStoxx50 equity index. In this case, the calibration of Student t copula will be super slow.

This indicates that the t distribution is more promising than the commonly used Student t copula when modeling the default correlations. Skewed (Student) t distribution has potential to become a powerful tool for quantitative analyst doing rich-cheap analysis of credit derivatives.

5.7 Conclusion

We show that Kendall's tau remains under monotone transformations so that both copula and distribution can be used to model the correlation of default times by the correlation of underlying equities.

We follow Rukowski's (1999) single name credit risk modeling and Schöbucher's (2003) portfolio credit risk modeling to price the basket credit default swaps.

The Student t copula is widely used in the pricing of basket credit default swaps for its lower tail dependence. However, we need to specify the marginal distributions first and calibrate the marginal distributions and copula separately. In addition, there is no good method to calibrate the degree of freedom ν. The calibration is very slow.

We use a fast EM algorithm for Student t distribution and skewed t distribution. All the parameters are calibrated together. To our knowledge, we are the first to introduce distribution to price basket credit default swaps.

The Student t copula leads to higher default probabilities and spread price of basket credit default swaps for LTD and lower default probabilities and spread price for FTD than the Gaussian copula.

Both the Student t distribution and the skewed t distribution lead to higher spread prices of basket credit default swaps for LTD and lower spread prices for FTD than the Student t copula.

APPENDIX A

Proof of equation 2.4.36, 2.4.40 and 2.4.42

We need following Bessel derivative formulas to prove those equations.

Derivatives of Bessel function. (Abramowitz and Stegun(1968) and Barndorff-Nielsen and Blæsild(1981)).

$$(\log K_\lambda(x))' = \frac{\lambda}{x} - \frac{K_{\lambda+1}(x)}{K_\lambda(x)}, \tag{5.7.1}$$

and

$$(\log K_\lambda(x))' = -\frac{\lambda}{x} - \frac{K_{\lambda-1}(x)}{K_\lambda(x)}, \tag{5.7.2}$$

where the first derivative equation is used to show equations 2.4.36 and 2.4.40 and the second derivative equation is used to show equation 2.4.42.

After we replace w_i^{-1}, w_i and $\log(w_i)$ by their conditional expectations $\delta_i^{[\cdot]}$, $\eta_i^{[\cdot]}$ and $\xi_i^{[\cdot]}$, the log likelihood L_2, i.e. equation 2.4.5 can be rewritten as

$$L_2(\lambda,\chi,\psi) = (\lambda-1)n\bar{\xi} - \frac{\chi}{2}n\bar{\delta} - \frac{\psi}{2}n\bar{\eta} - \frac{n\lambda}{2}\log\chi + \frac{n\lambda}{2}\log\psi - n\log\left(2K_\lambda(\sqrt{\chi\psi})\right).$$

In the maximization of L_2, we usually set λ to be a constant.

By setting $\frac{\partial L_2}{\partial \chi} = 0$[§], we can get

$$\bar{\delta} + \frac{2\lambda}{\chi} - \sqrt{\frac{\psi}{\chi}}\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} = 0. \tag{5.7.3}$$

By setting $\frac{\partial L_2}{\partial \psi} = 0$[¶], we can get

$$\chi = \frac{\theta\bar{\eta}K_\lambda(\theta)}{K_{\lambda+1}(\theta)}. \tag{5.7.4}$$

By plugging equation 5.7.4 back to equation 5.7.3, we can get equation 2.4.36.

If we set χ to be a constant, from equation 5.7.4, we can get equation 2.4.40.

If we set ψ to be a constant,by setting $\frac{\partial L_2}{\partial \chi} = 0$[‖], we can get equation 2.4.42.

[§]We use equation 5.7.1.
[¶]We use equation 5.7.1.
[‖]We use equation 5.7.2.

REFERENCES

Aas, K. and Hobæk Haff, I.(2005a). The generalized hyperbolic skew student's t-distribution. *Submitted to Journal of the Royal Statistical Society, C.*

Aas, K. and Hobæk Haff , I.(2005b). NIG and skew student's t: two special cases of the generalized hyperbolic distribution. Norwegian Computing Center, Oslo, Norway.

Aas, K., Dimakos, X. and Hobæk Haff , I.(2005c). Risk estimation using the multivariate normal inverse Gaussian distribution. *Submitted to the Journal of Risk*.

Abramowitz,M and Stegun, I.(1968). *Handbook of Mathematical Functions.* New York: Dover Publ.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Math. Finance*, 9:203-228.

Azzalini, A. and Capitanio, A.(1998). Statistical applications of the multivariate skew-normal distribution. *J. Roy. Statist. Soc.,* Series B, vol. 61, no.3.

Barbachan, Farias and Ornelas(2003). Analyzing the use of generalized hyperbolic distributions to value at risk calculations. Finance Lab, Ibmec Business School (Sao Paulo). Series: Finance Lab Working Papers, number flwp_58.

Barndorff-Nielsen, O.E. (1977). Exponentially decreasing distributions for the logarithm of the particle size. *Proceedings of the Royal Society.* London. Series A. Mathematical and Physical Sciences, 353:401-419.

Barndorff-Nielsen, O.E. and Blæsid (1981). Hyperbolic distributions and ramifications: Contributions to theory and application. In C. Taillie, G. Patil, and B. Baldessari(Eds.), *Statistical Distributions in Scientific Work,* Vol. 4, Dordrecht: Reidel.

Barndorff-Nielsen, O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modeling. *Scand. J. Statist.* 24:1-13, 1997.

Barndorff-Nielsen, O.E. and Stelzer, R (2004). Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy-Processes. Department of Mathematical Sciences, University of Aarhus, Denmark and Centre for Mathematical Sciences, Munich University of Technology, Germany.

Belkacem, L.(1997). How to select optimal portfolio in α-stable markets. Institut National De Recherche En Informatique ET EN Automatique.

Blæsild, P. (1981). The two-dimensional hyperbolic distribution and related distributions, with an application to johannsen's bean data. *Biometrika*, 68(1):251-263.

Blæsild, P. and Sørensen, M.(1992). 'hyp'- a computer program for analyzing data by means of the hyperbolic distribution. Research Report 248, Department of Theoretical Statistics, University of Aarhus.

Bradley, B. and Taqqu, M.(2001). Financial risk and heavy tails. In *Heavy-Tailed Distributions in Finance,* Rachev, S., North Holland.

Brigo, D., and Mercurio, F.(2005). *Interest Rate Models -Theory and Practice.* Springer Finance.

Cherubini, U., Luciano, E. and Vecchiato, W.(2004). *Copula Methods in Finance.* John Wiley & Sons, Ltd.

Chen, Y., Hardle, W., and Jeong, S.(2004). Nonparametric risk management with generalized hyperbolic distributions. Center for Applied Statistics and Economics, Humboldt-Universitat zu Berlin, and Institut de Statisque, Universite Catholique de Louvain.

Demarta, Stefano (2002). Diploma thesis: Multivariate extreme value theory and copulas. Federal Institute of Techonogy, ETH Zentrum, Zurich.

Demarta, S. and McNeil, A. (2005). The t copula and related copulas. *International Statistical Review, (to appear 2005).*

Dempster, A.P., Laird, N.M. and Rubin, D.B.(1977). Maximum likelihood from incomplete data via the EM algorithm(with discussion). *J. Roy. Statist. Soc*, Ser B 39 1-38.

Di Clemente, A. and Romano, C. (2003a). Beyond Markowitz: building the optimal portfolio using non-elliptical asset return distributions. working paper, CIDEM, University of Rome, ′La Sapienza′.

Di Clemente, A. and Romano, C. (2003b). A copula-extreme value theory approach for modeling operational risk. working paper. Department of Economics Theory and Quantitative Methods for the Political Choices, University of Rome, Italy. Risk Management, Control and Capital Allocation Dept, Capitalia Bank Holding, Italy.

Di Clemente, A. and Romano, C. (2003c). Measuring portfolio value-at-risk by a copula EVT Based approach. working paper. Department of Economics Theory and Quantitative Methods for the Political Choices, University of Rome, Italy. Risk Management, Control and Capital Allocation Dept, Capitalia Bank Holding, Italy.

Di Clemente, A. and Romano, C. (2004). Measure and optimizing portfolio credit risk: a copula-based approach. working paper. Department of Economics Theory and Quantitative Methods for the Political Choices, University of Rome, Italy. Risk Management, Control and Capital Allocation Dept, Capitalia Bank Holding, Italy.

Duffie, Darrel and Singleton, Kenneth (2003). *Credit Risk: Pricing, Measurement, and Management.* Princeton University Press.

Durrleman, V., Nikeghbali, A., Riboulet, G. and Roncalli, T. (2000). Copulas for finance: a reading guide and some applications. groupe de recherche operationnelle, Credit lyonnais.

Eberlein, Ernst and Keller, Ulrich (1995). Hyperbolic distributions in finance. *Bernoulli* 1: 281-299.

Eberlein, Ernst and Hammerstein, Ernst (2003). Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes. University of Freiburg. Nr.80.

Embrechts, P., Kluppelberg, C. and Mikosch, T.(1997). Modeling extremal events for insurance and finance. Springer.

Embrechts, P., McNeil, A., and Straumann(2001). Correlation and dependency in risk management: properties and pitfalls. In *Risk Managemenmt: Value at Risk and Beyond,* ed. by M. Dempster, and H. Moffatt. Cambridge University Press.

Embrechts, P., Lindskog, F. and McNeil, A.(2001). Modeling dependence with copulas and applications to risk management. Department of mathematics, ETHZ, CH-8092, Zürich, Switzerland.

Fajardo, J. and Farias, A.(2003). Generalized hyperbolic distributions and Brazilian data. Ibmec Business School. FinanceLab Working Paper, FLWP-2003-10.

Fang, K.T., Kotz, S. and Ng, K.W.(1990). *Symmetric Mulrivariate and Related Distributions.* Chapman & Hall, London.

Frahm, G., Junker, M. and Schmidt, R.(2004). Estimating the tail-dependence coefficient: properties and pitfalls. Research Center Caesar, Research Center Caesar and London School of Economics.

Frees, E. and Valdez, E.(1997). Understanding relationships using copulas. 32nd Actual Research Conference, Aug 6-8, 1997, University of Calgary, Calgary, Alberta, Canada.

Galiani, Stefano (2003). Copula functions and their applications in pricing and risk managing multiname credit derivative products, Department of Mathematics, King's Colledge, Longdon, UK. Master Thesis.

Gencay, R., Selcuk, F., and Ulugulyagcl, A. (2001). EVIM: A software package for extreme value analysis in Matlab. *Studies in Nonlinear Dynamics and Econometrics*, Oct 2001, 5(3):213-239.

Georges, P., Lamy, A., Nicolas, E., Quibel, G. and Roncalli, T.(2001). Multivariate survival modelling: a unified approach with copulas. *Credit Lyonnais.*

Giesecke, Kay (2003a). Correlated default with incomplete information, *to appear in Journal of Banking and Finance.*

Giesecke, Kay (2003b). A simple exponential model for dependent defaults, *Journal of Fixed Income* **13**(3), 74–83, 2003.

Hull, John(2002). *Options, Futures and Other Derivatives.* 5th edition. Prentice Hall.

Joe, H. (1996). Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, 57 240-265(1996). Article No. 0032.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Monographs on statistics and applied probability, 73. Chapmann & Hall, London.

Jørgensen, B(1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution,* Lecture notes in statistics, Springer, Heidelberg.

Jouanin J.-F., Rapuch, G., Roncalli, T. (2001). Modeling dependence for credit derivatives with copulas, Credit Lyonnais, France.

Kupiec, P.(1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, Winter, 73-84.

Li, David(1999). On default correlation: A copula function approach, working paper, The RiskMetrics Group.

Lindskog, McNeil, and Schmock(2003). Kendall's tau for elliptical distributions. In *Credit Risk-measurement, Evaluation and mManagement,* Bol, Nakhaeizade et al., eds. Physica-Verlag, Heidelberg.

Liu, C. and Rubin, D.(1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica,* 5(1995), 19-39.

Markowitz, H.(1952). Portfolio selection. *Journal of Finance*, 7:77-91, 1952.

Masala, G., Menzietti, M., and Micocci, M. (2004). Optimisation of conditional-VaR in an actuarial model for credit risk assuming a student copula denpendence structure. University of Cagliari, University of Rome, and University of Cagliari.

Mashal R. and Naldi M.(2002). Pricing multiname credit derivatives: heavy tailed hybrid approach. Working paper. Columbia university and Lehman Brothers Inc.

Marshall, A. and Olkin, I.(1988). Families of multivariate distributions. *Journal of American Statiscal Association,* Sept 1988, Vol. 83, No.403.

McNeil, A., Frey, R. and Embrechts, P.(2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* The Draft of June 21, 2004. Will be published by Princeton University Press(2005).

McNeil, A. (1999). Extreme value theory for risk managers. Department of Mathematik, ETH Zentrum, CH-8092 Zürich.

Meneguzzo, D. and Vecchiato, W.(2002). Copula sensitivity in collateralized debt obligations and basket default swaps pricing and risk monitoring. Draft.

Nelsen, Roger (1999). *An Introduction to Copulas.* Lectures notes in statistics, 139, Springer Verlag.

Nolan,John(2005). Stable distributions: models for heavy tailed data. Math/Stat Department, American University.

Prause, K.(1999). The Generalized hyperbolic model: estimation, financial derivatives, and risk measures. Tese de Doutorado, University of Freiburg.

Protassov, Rostislav(2004). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ. *Statistics and Computing.* Vol 14, issue 1.

Rachev, S. and Mittnik, S.(2000). *Stable Paretian Models in Finance.* John Wiley & Sons, Ltd.

Rockafellar, R.T. and Uryasev, S.(1999). Optimization of conditional value at risk. Research report 99-4, Center for Applied Optimization, University of Florida.

Romano, C. (2002a). Calibrating and simulating copula functions: an application to the Italian stock market. Risk Management Function, Capitalia, Viale U. Tupini, 180, 00144-Rome, Italy.

Romano, C. (2002b). Applying copula function to risk management. Working paper. Risk Management Function, Capitalia, Viale U. Tupini, 180, 00144- Rome, Italy.

Rubin, D.B.(1983). Iteratively reweighted least squares. *Encyclopedia of Statistical Sciences*, Vol. 4, 272-275 John Wiley, New York.

Rogge, Ebbe and Schönbucher, P.J.(2003). Modeling dynamic portfolio credit risk. Department of Mathematics, Imperial College and ABN AMRO Bank, London and Department of Mathematics, ETH Zürich, Zürich. *Working Paper.*

Rutkowski(1999). On models of default risk by R. Elliott, M. Jeanblanc and M. Yor. Warsaw University of Techonology.

Schmidt, R.(2002). Tail dependence for elliptically contoured distributions. *Math Meth Oper Res*, 2002, 55:301-327.

Schmidt R.(2003a). Dependencies of extreme events in finance: modeling, statistics, and data analysis. Dissertation. ULM. 2003.

Schmidt R.(2003b). Tail dependence. In *Statistical tools in Finance and Insurance*, P. Cizek, W. Hardle, and R. Weron. Xplore e-books, Springer Verlag, 2003.

Schmidt R.(2003c). Credit risk modeling and estimation via elliptical copulae. in *Credit Risk-Measurement, Evaluation, and Management*, Physica-Verlag Heidelberg, 267-289.

Schönbucher, P.J. and D. Schubert (2001). Copula-dependent default risk in intensity models, Department of Statistics, Bonn University. *Working Paper.*

Schönbucher, P.J.(2003). *Credit Derivatives Pricing Models: Models, Pricing and Implementation.* John Wiley & Sons.

Schoutens, W. (2003). *Levy Processes in Finance: Pricing Financial Derivatives.* John Wiley & Sons, Ltd.

Sklar, A(1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist.* Univ. Paris 8, 229-231.

BIOGRAPHICAL SKETCH

Wenbo Hu

Wenbo Hu was born in Sept, 1976, in Jingdezhen, Jiangxi, P.R. China. In summer of 1998, he completed his Bachelor's degree in Statistics at Zhongshan University. Under the advisement of Prof. Qiansheng Cheng, he obtained his Master's degree in summer of 2001, from the Department of Financial Mathematics and Department of Informatics, School of Mathematical Sciences at Peking University. He obtained the master's degree in Financial Mathematics at Florida State University in summer of 2003 under the advisement of Prof. Bettye Case. He enrolled in the Doctoral program at Florida State University in summer of 2003 under the advisement of Prof. Alec Kercheval.

Hu's research interests include Generalized Hyperbolic Distributions, Copulas, Portfolio Management, Risk Management, (Portfolio) Credit Risk, Fixed Income, Option Pricing, Automatic Trading System Design, Pattern Recognition, Artificial Neural Networks, GARCH Models, Global Optimization Methods. Hu also has interest in actuarial science and has passed a number of exams of Society of Actuaries.

He worked for United Securities in 2000 to design automatic equity trading system. He is working for Bell Trading to design treasuries, equities, energies, and currencies automatic trading system.

Hu lives in Breckenridge, Colorado now.